



US009334517B2

(12) **United States Patent**  
**Lee et al.**

(10) **Patent No.:** **US 9,334,517 B2**  
 (45) **Date of Patent:** **May 10, 2016**

(54) **ENDOGLUCANASE HAVING ENHANCED THERMOSTABILITY AND ACTIVITY**

(71) Applicant: **CALIFORNIA INSTITUTE OF TECHNOLOGY**, Pasadena, CA (US)

(72) Inventors: **Toni M. Lee**, Orange, CA (US); **Stephen L. Mayo**, Pasadena, CA (US); **Frances H. Arnold**, La Canada, CA (US); **Devin Trudeau**, Pasadena, CA (US)

(73) Assignee: **California Institute of Technology**, Pasadena, CA (US)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **14/247,221**

(22) Filed: **Apr. 7, 2014**

(65) **Prior Publication Data**

US 2014/0308713 A1 Oct. 16, 2014

#### **Related U.S. Application Data**

(60) Provisional application No. 61/809,144, filed on Apr. 5, 2013, provisional application No. 61/822,489, filed on May 13, 2013.

(51) **Int. Cl.**  
**C12P 19/14** (2006.01)  
**C12N 9/42** (2006.01)

(52) **U.S. Cl.**  
 CPC ..... **C12P 19/14** (2013.01); **C12N 9/2437** (2013.01); **C12Y 302/01004** (2013.01); **C12Y 302/01091** (2013.01); **Y02E 50/343** (2013.01); **Y02P 20/52** (2015.11)

(58) **Field of Classification Search**  
 None  
 See application file for complete search history.

(56) **References Cited**

#### **FOREIGN PATENT DOCUMENTS**

WO WO 2011/109905 \* 9/2011  
 WO WO 2012/106824 \* 8/2012

#### **OTHER PUBLICATIONS**

Allen, Benjamin D. et al.; "Dramatic Performance Enhancements for the FASTER Optimization Algorithm"; 2006; Journal of Computational Chemistry; vol. 27; No. 10; pp. 1071-1075.  
 Altschul, Stephen F. et al.; "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs"; 1997; Nucleic Acids Research; 25; pp. 3389-3402.  
 Celik, Eda et al.; "Production of recombinant proteins by yeast cells"; 2012; Biotechnology Advances; 30; pp. 1108-1118.  
 Chica, Roberto A. et al.; "Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries"; Nov. 23, 2010; PNAS; vol. 107; No. 47; pp. 20257-20262.  
 Dahiyat, Bassil I. et al.; "Probing the role of packing specificity in protein design"; Sep. 1997; PNAS; 94; pp. 10172-10177.

Dahiyat, Bassil I. et al.; "Automated design of the surface positions of protein helices"; 1997; Protein Science; 6; pp. 1333-1337.

Desmet, Johan et al.; "Fast and Accurate Side-Chain Topology and Energy Refinement (FASTER) as a New Method for Protein Structure Optimization"; 2002; Proteins: Structure, Function, and Genetics; 48; pp. 31-43.

Dombkowski, Alan A.; "Disulfide by Design™: a computational method for the rational design of disulfide bonds in proteins"; 2003; Bioinformatics; vol. 19; No. 14; pp. 1852-1853.

Edgar, Robert C.; "MUSCLE: multiple sequence alignment with high accuracy and high throughput"; Mar. 19, 2004; Nucleic acids research; vol. 32; No. 5; pp. 1792-1797.

Gasteiger, Elisabeth et al.; "Protein Identification and Analysis Tools on the ExPASy Server"; 2005; The Proteomics Protocols Handbook; pp. 571-607.

Gibson, Daniel G.; "Enzymatic Assembly of Overlapping DNA Fragments"; 2011; Methods in Enzymology; Chapter 15; vol. 498; pp. 349-361.

Gibson, Daniel G. et al.; "Enzymatic assembly of DNA molecules up to several hundred kilobases"; May 2009; Nature Methods; vol. 6; No. 5; pp. 343-345.

Green III, Frederick et al.; "Adaptation of the Nelson-Somogyi Reducing-Sugar Assay to a Microassay Using Microtiter Plates"; 1989; Analytical Biochemistry; 182; pp. 197-199.

Heinzelman, Pete et al.; "A family of thermostable fungal cellulases created by structure-guided recombination"; Apr. 7, 2009; PNAS; vol. 106; No. 14; pp. 5610-5615.

Komor, Russell S. et al.; "Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods"; 2012; Protein Engineering, Design & Selection; vol. 25; No. 12; pp. 827-833.

Lee, Toni M. et al.; "A structural study of Hypocrea jecorina Cel5A"; Sep. 2011; Protein Science; vol. 20; pp. 1935-1940.

Marshall, Shannon A. et al.; "Electrostatics Significantly Affect the Stability of Designed Homeodomain Variants"; 2002; J. Mol. Biol.; 316; pp. 189-199.

Matlab; "Statistics and Machine Learning Toolbox Release Notes"; 2011, The MathWorks, Inc.; 138pp.

Mayo, Stephen L. et al.; "DREIDING: A Generic Force Field for Molecular Simulations"; 1990; J. Phys. Chem.; 94; pp. 8897-8909.

Metropolis, Nicholas et al.; "Equation of State Calculations by Fast Computing Machines"; Jun. 1953; The Journal of Chemical Physics; vol. 21; No. 6; pp. 1087-1092.

Murray, Patrick et al.; "Expression in Trichoderma reesei and characterisation of a thermostable family 3  $\beta$ -glucosidase from the moderately thermophilic fungus Talaromyces emersonii"; 2004; Protein Expression & Purification; 38; pp. 248-257.

Nelson, Norton; "A Photometric Adaptation of the Somogyi Method for the Determination of Glucose"; Feb. 3, 1944; J. Biol. Chem.; 153; pp. 375-379.

Park, James T. et al.; "A submicrodetermination of glucose"; Jun. 23, 1949; J. Biol. Chem.; 181; pp. 149-151.

(Continued)

*Primary Examiner* — Nashaat Nashed

(74) *Attorney, Agent, or Firm* — Lewis Roca Rothgerber Christie LLP

(57) **ABSTRACT**

A variant Cel5a endoglucanase has increased thermostability, increased enzymatic activity and/or increased expression in a host, relative to wild type Cel5a. The improved variant Cel5a endoglucanase may be used to hydrolyze more cellulose at a higher temperature for a more efficient and cost-effective production of biofuels as compared to wild type Cel5a.

**17 Claims, 37 Drawing Sheets**

(56)

**References Cited**

## OTHER PUBLICATIONS

- Qin, Yuqi et al.; "Purification and characterization of recombinant endoglucanase of *Trichoderma reesei* expressed in *Saccharomyces cerevisiae* with higher glycosylation and stability"; 2008; Protein Expression & Purification; 58; pp. 162-167.
- Rohl, Carol A. et al.; "[4] Protein Structure Prediction Using Rosetta"; 2004; Methods in Enzymology; vol. 383; pp. 66-93.
- Romero, Philip A. et al.; "Schema-Designed Variants of Human Arginase I and II Reveal Sequence Elements Important to Stability and Catalysis"; 2012; ACS Synthetic Biology; 1; pp. 221-228.
- Schiestl, Robert H. et al. "Introducing DNA into yeast by transformation," 1993; Methods: Companion to Methods in Enzymology; 5; pp. 79-85.
- Schymkowitz, Joost et al.; "The FoldX web server: an online force field"; 2005; Nucleic Acids Research; vol. 33; pp. W382-W388.
- Smith, Matthew A. et al.; "Hypocrea jecorina Cellobiohydrolase I Stabilizing Mutations Identified Using Noncontiguous Recombination"; 2013; ACS Synthetic Biology; 2; pp. 690-696.
- Somogyi, Michael; "Notes on Sugar Determination"; 1951; J. Biol. Chem.; 195; pp. 19-23.
- Su, Xiaoyun et al.; "Heterologous Gene Expression in Filamentous Fungi"; 2012; Advances in Applied Microbiology; vol. 81; Chapter 1; pp. 1-61.
- Sullivan, Brandon J. et al.; "Stabilizing Proteins from Sequence Statistics: The Interplay of Conservation and Correlation in Triosephosphate Isomerase Stability"; 2012; Journal of Molecular Biology; 420; pp. 384-399.
- Sun, Ye et al.; "Hydrolysis of lignocellulosic materials for ethanol production: a review"; 2002; Bioresource Technology; 83; pp. 1-11.
- Villalobos, Alan et al.; "Gene Designer: a synthetic biology tool for constructing artificial DNA segments"; 2006; BMC Bioinformatics; 7; 285; 8pp.
- Voigt, Christopher A. et al.; "Trading Accuracy for Speed: A Quantitative Comparison of Search Algorithms in Protein Sequence Design"; 2000; J. Mol. Biol.; 299; pp. 789-803.
- Wu, Indira et al.; "Engineered Thermostable Fungal Cel6A and Cel7A Cellobiohydrolases Hydrolyze Cellulose Efficiently at Elevated Temperatures"; 2013; Biotechnology and Bioengineering; vol. 110; No. 7; pp. 1874-1883.
- Ye, Xiaoduan et al.; "Hypergraph Model of Multi-Residue Interactions in Proteins: Sequentially-Constrained Partitioning Algorithms for Optimization of Site-Directed Protein Recombination"; 2007; Journal of Computational Biology; vol. 14; No. 6; pp. 777-790.

\* cited by examiner

FIG. 1

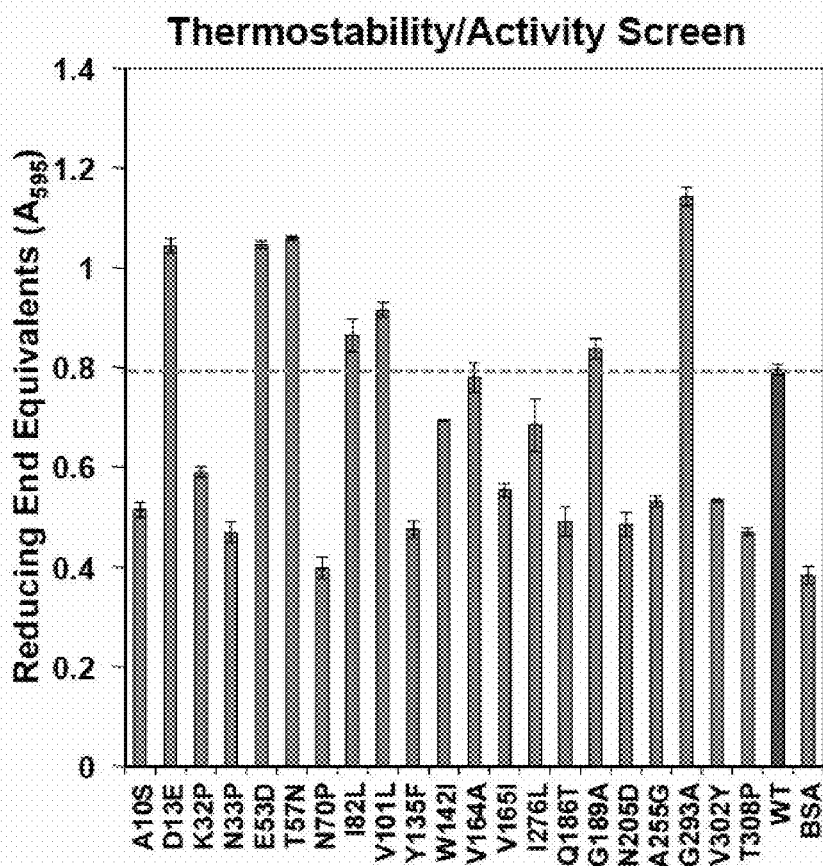


FIG. 2

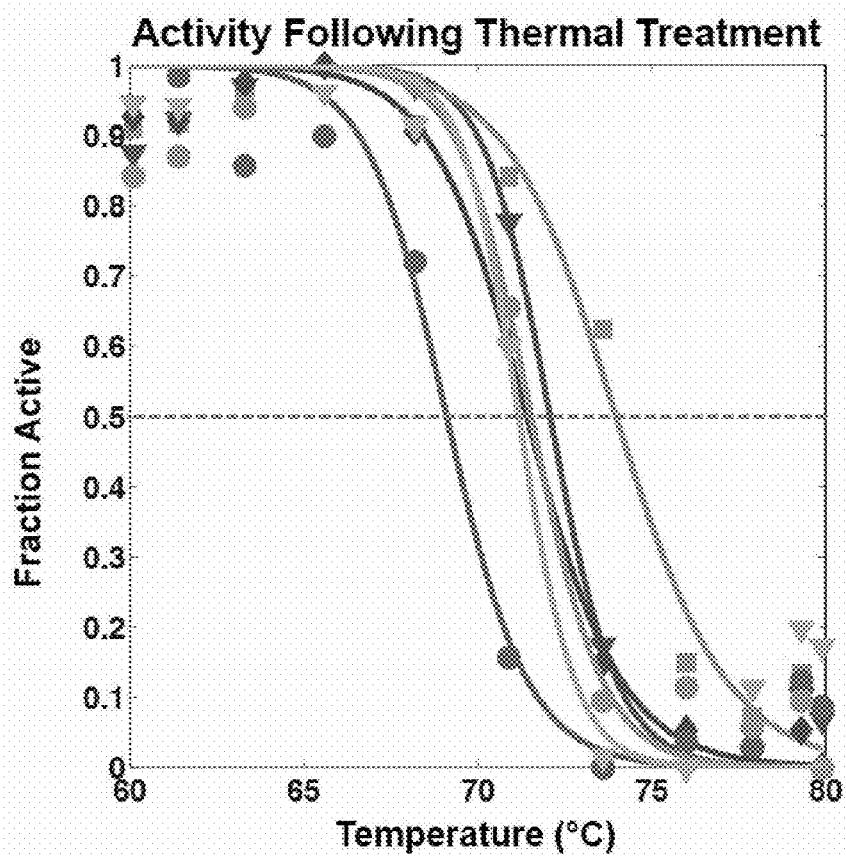


FIG. 3A

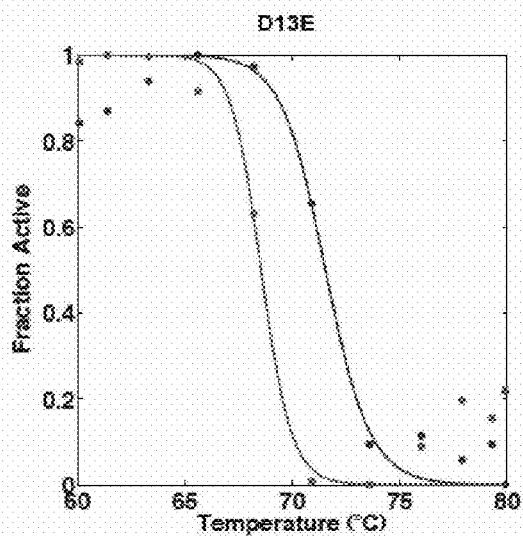


FIG. 3B

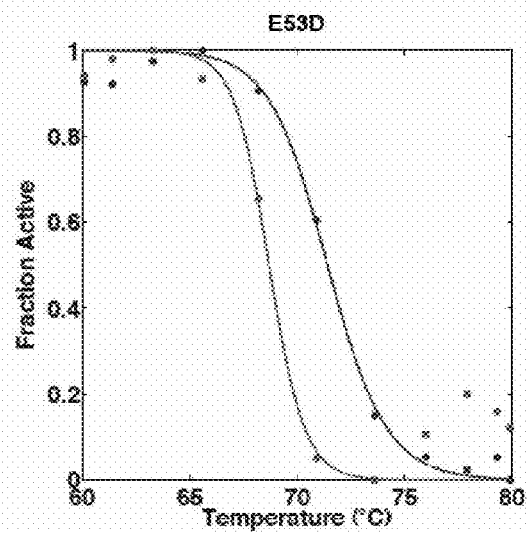


FIG. 3C

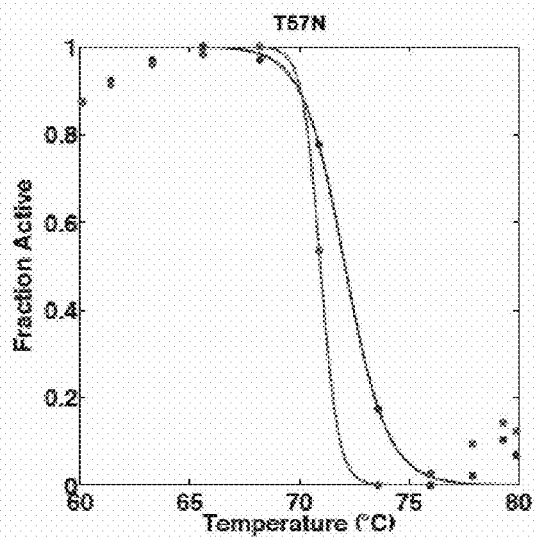


FIG. 3D

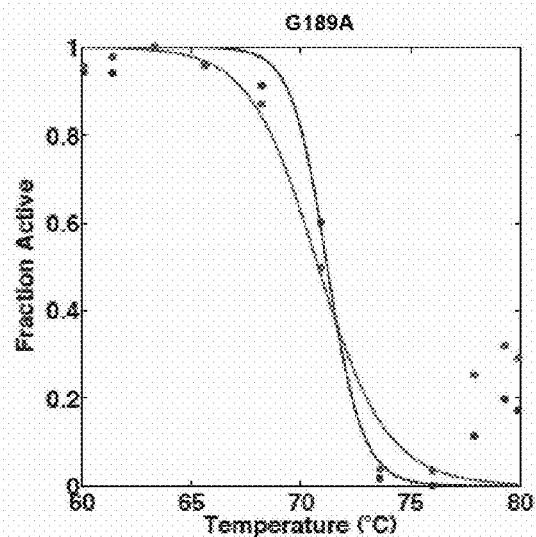


FIG. 3E

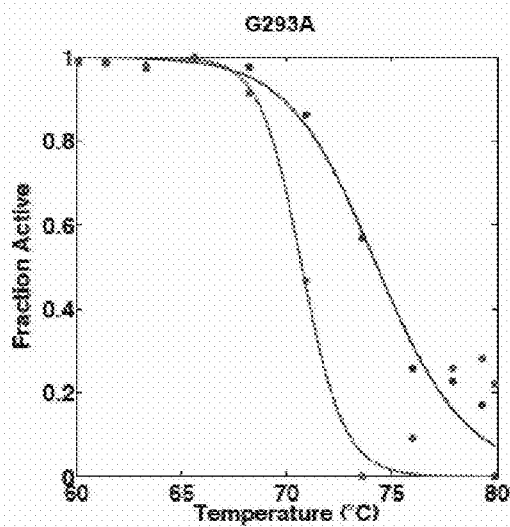


FIG. 3F

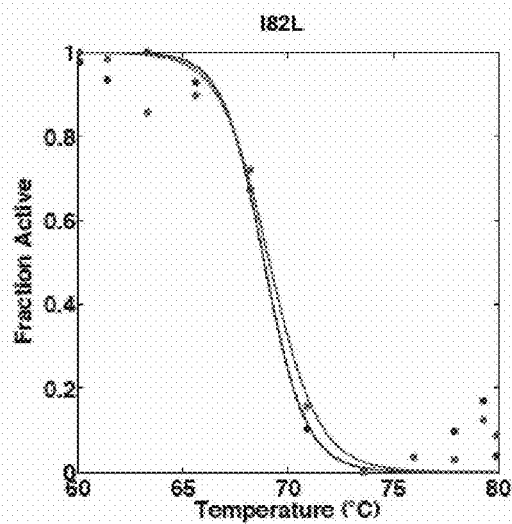


FIG. 3G

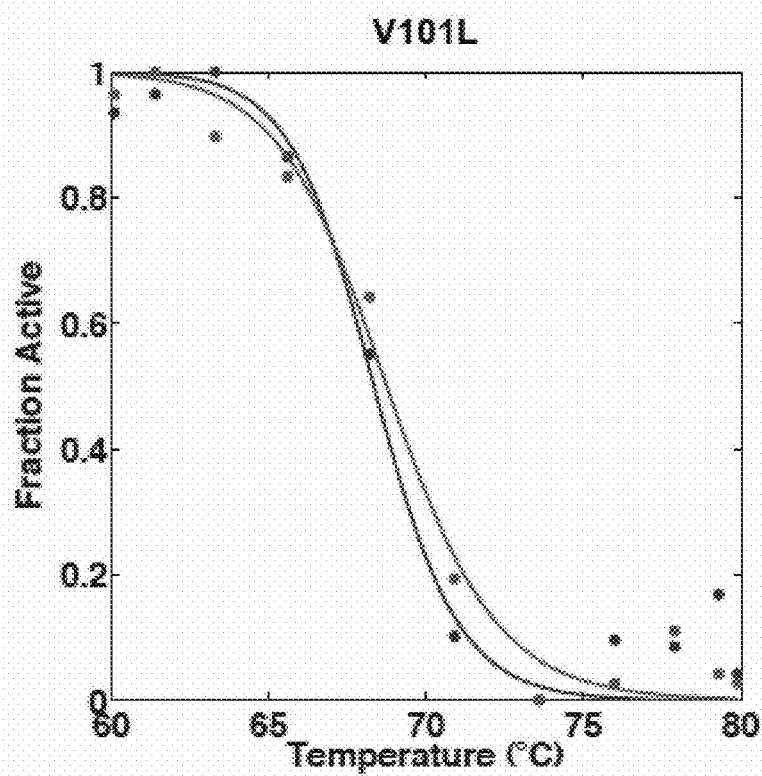




FIG. 4

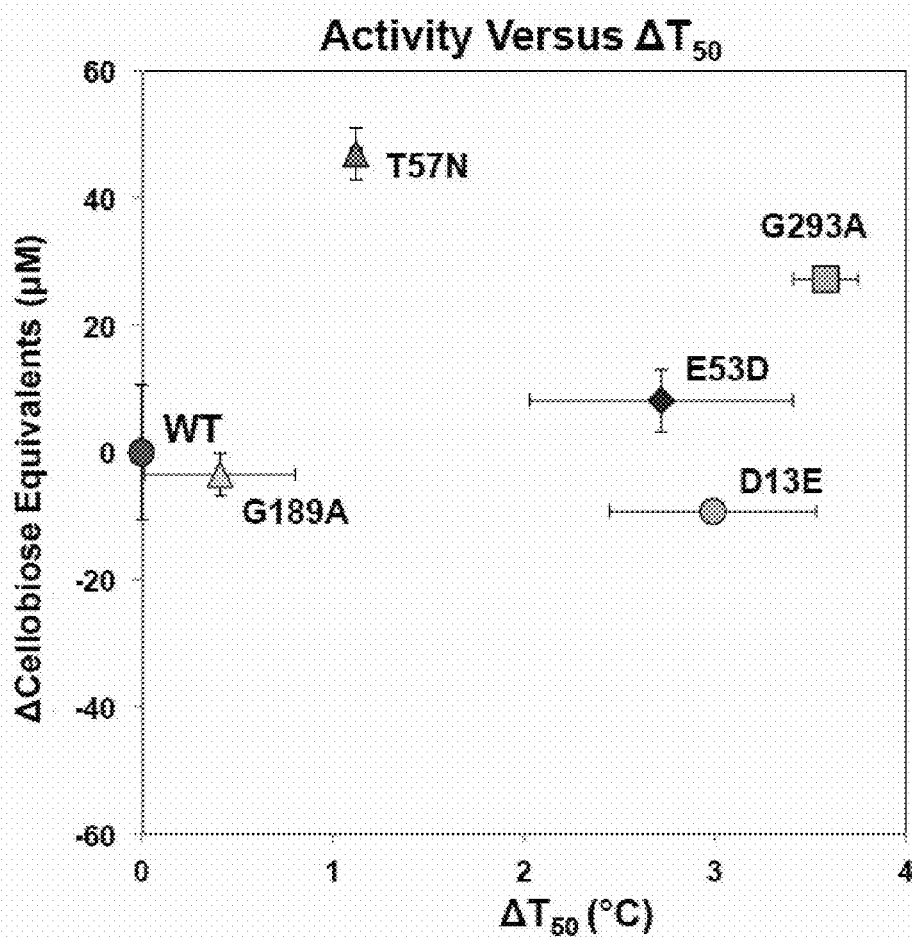


FIG. 5

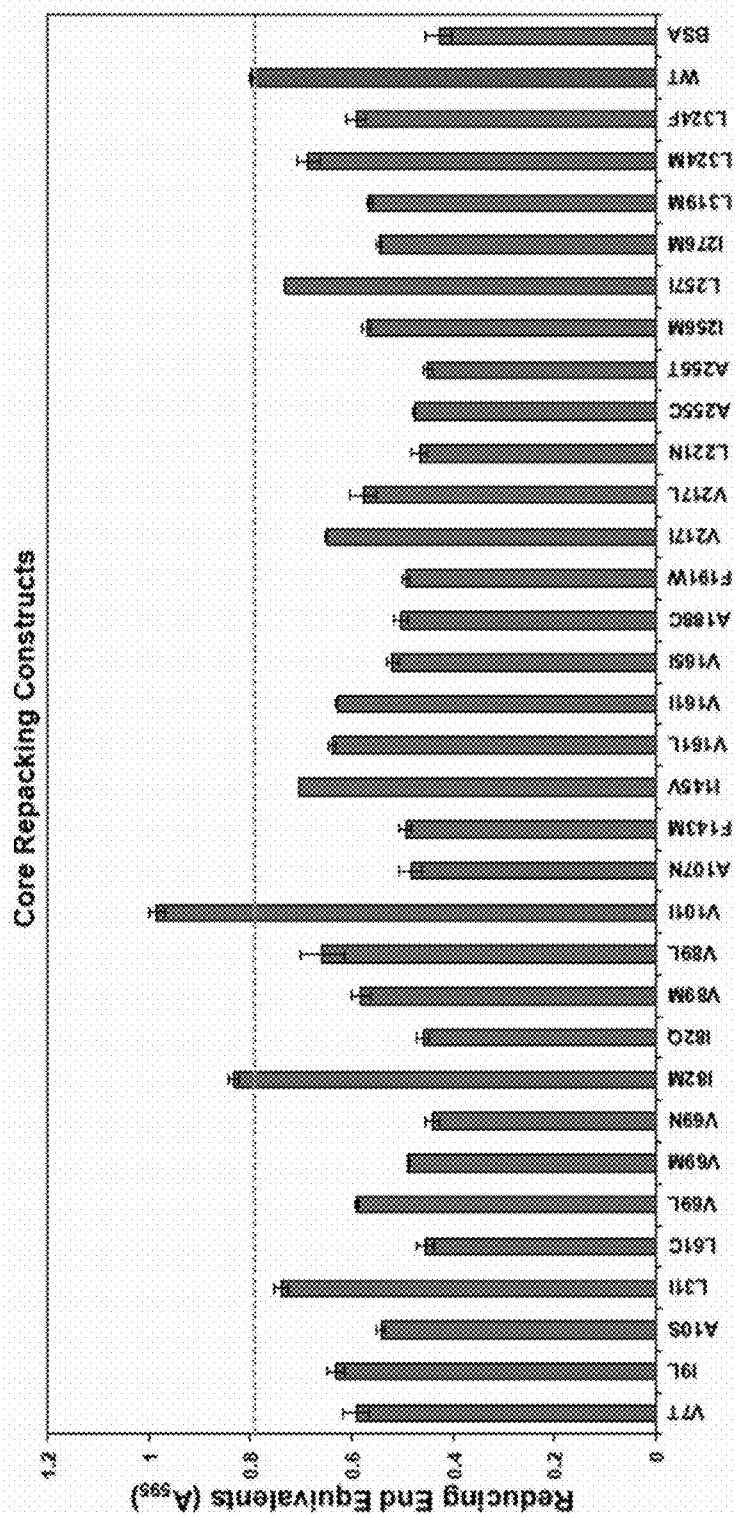


FIG. 6A

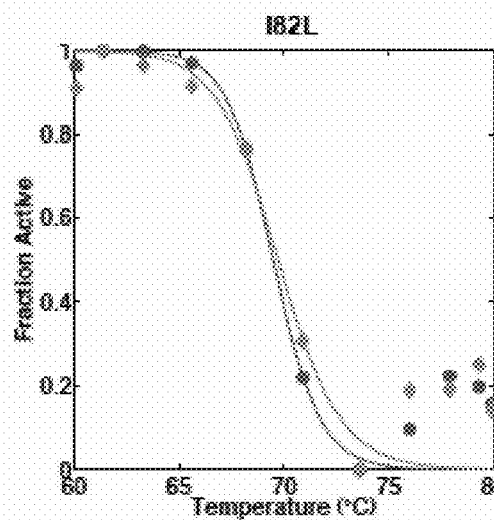


FIG. 6B

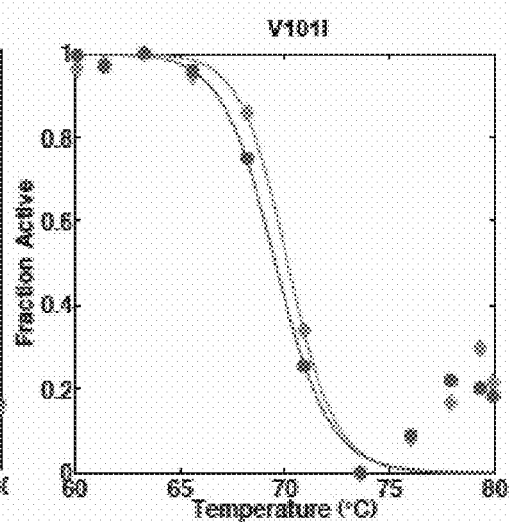


FIG. 7

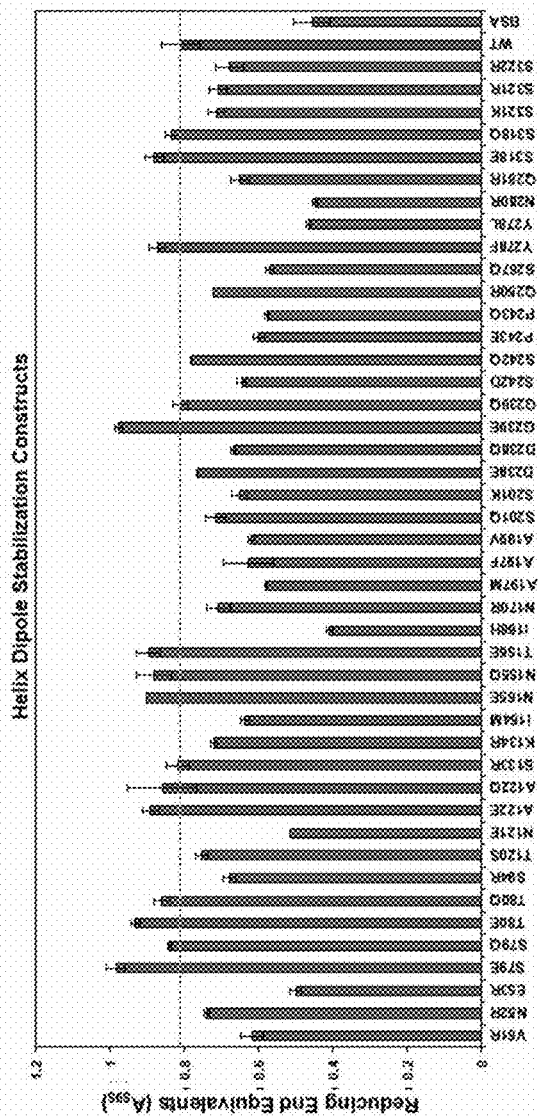


FIG. 8A

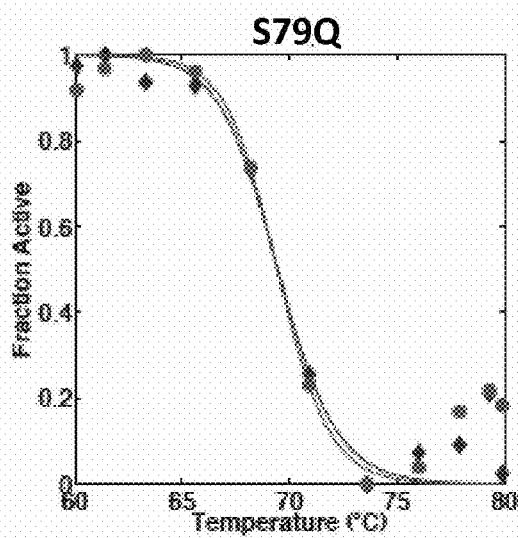


FIG. 8B

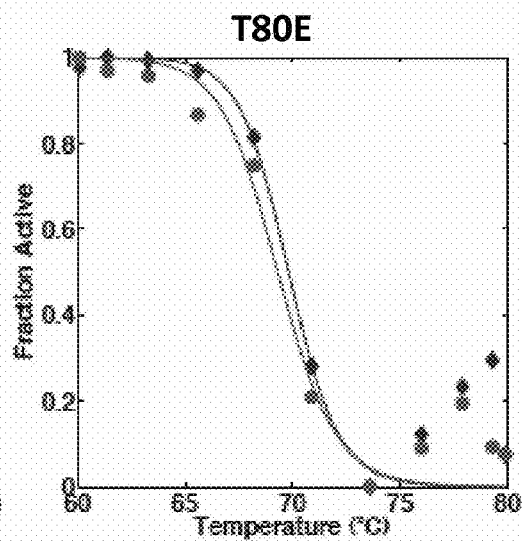


FIG. 8C

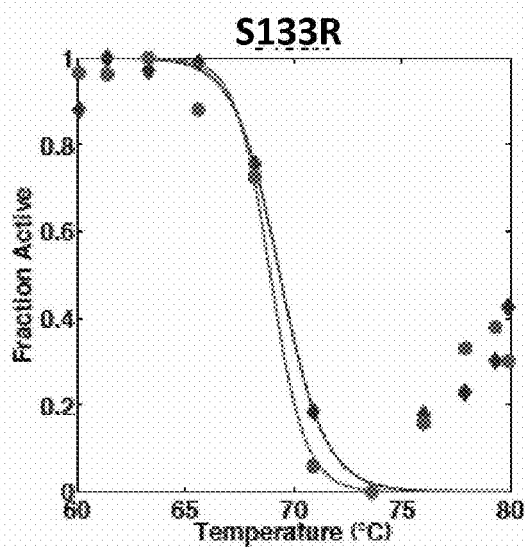


FIG. 8D

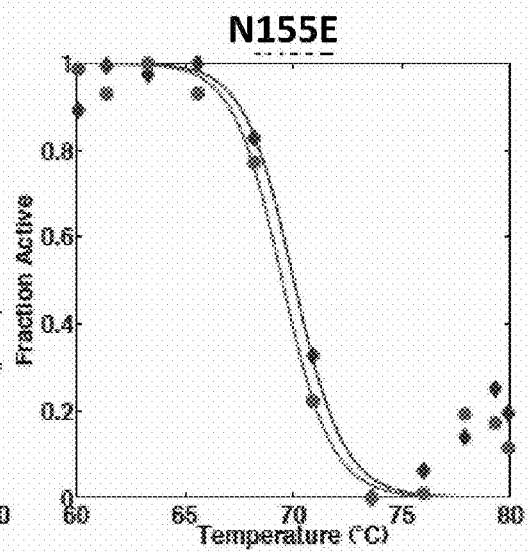


FIG. 8E

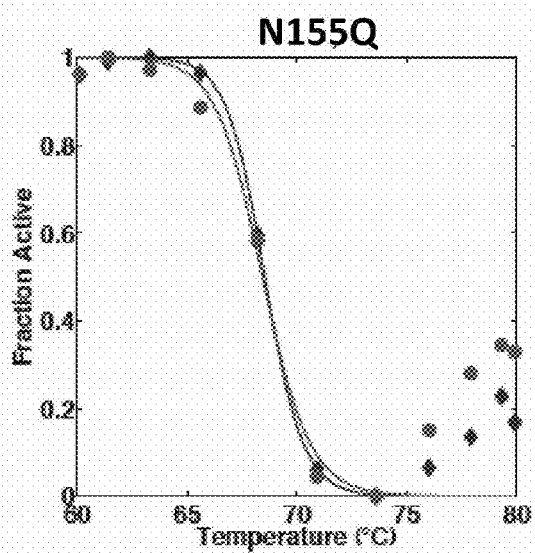


FIG. 8F

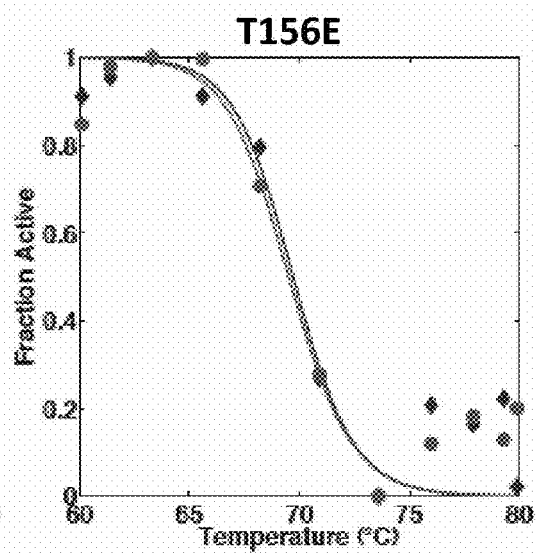


FIG. 8G

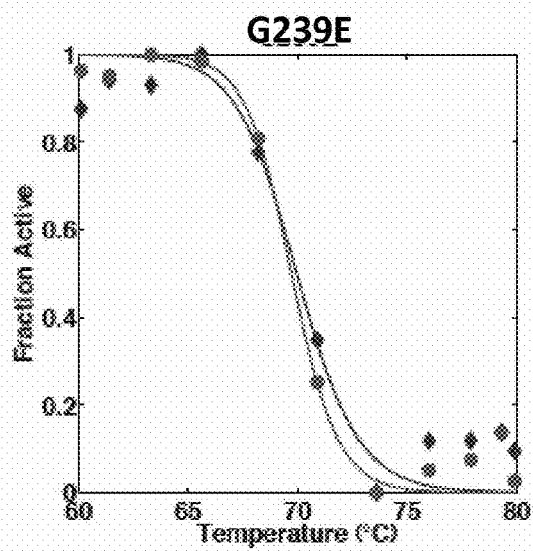


FIG. 8H

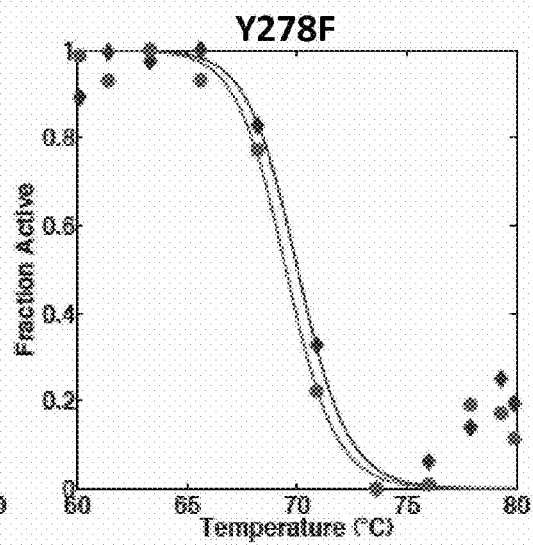




FIG. 8I

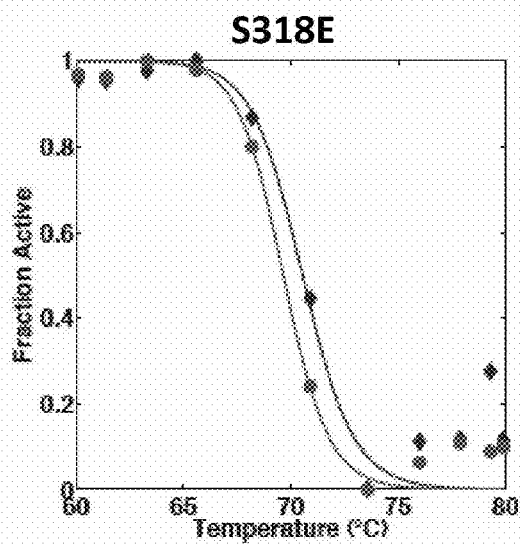


FIG. 8J

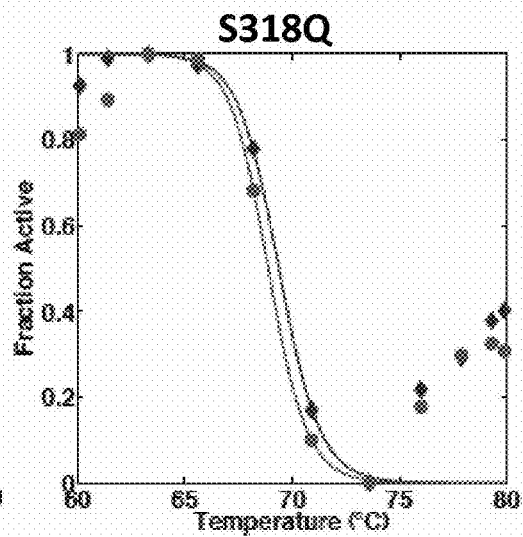


FIG. 8K

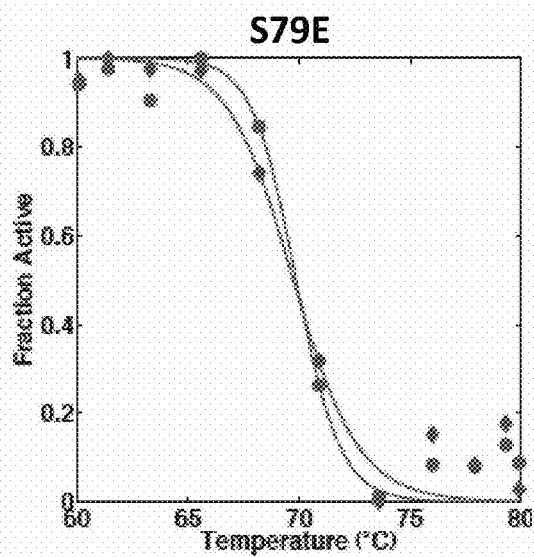


FIG. 8L

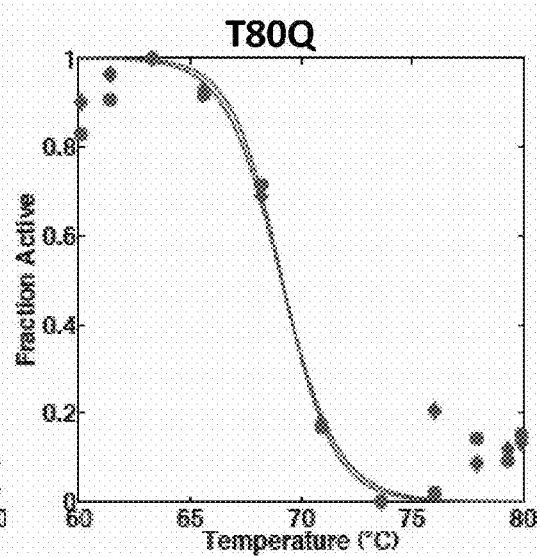


FIG. 8M

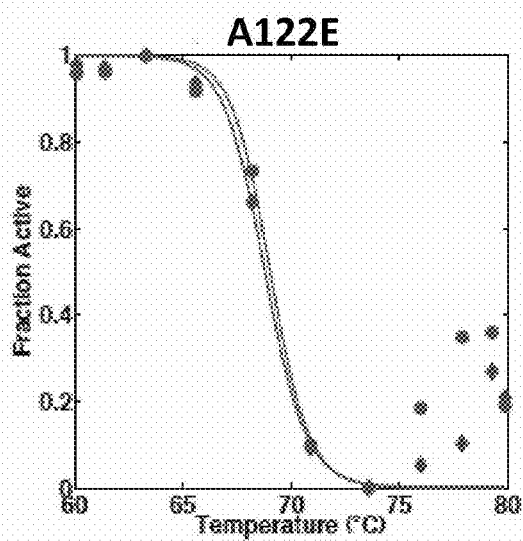


FIG. 8N

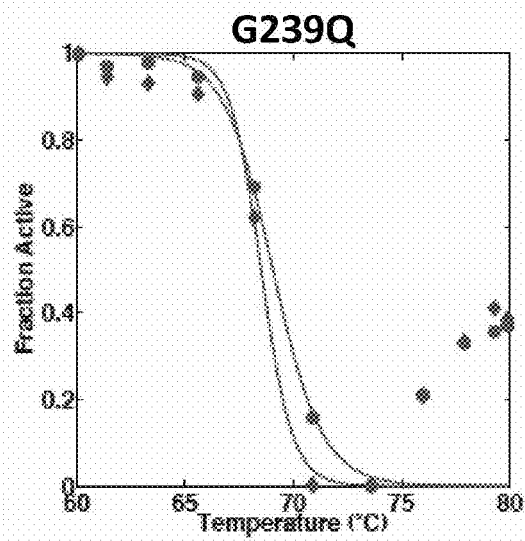


FIG. 9A

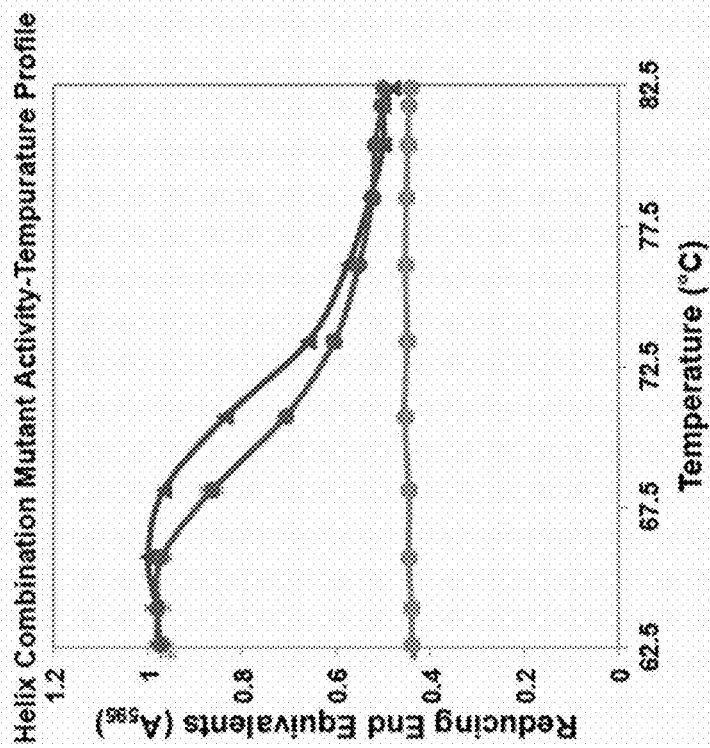


FIG. 9B

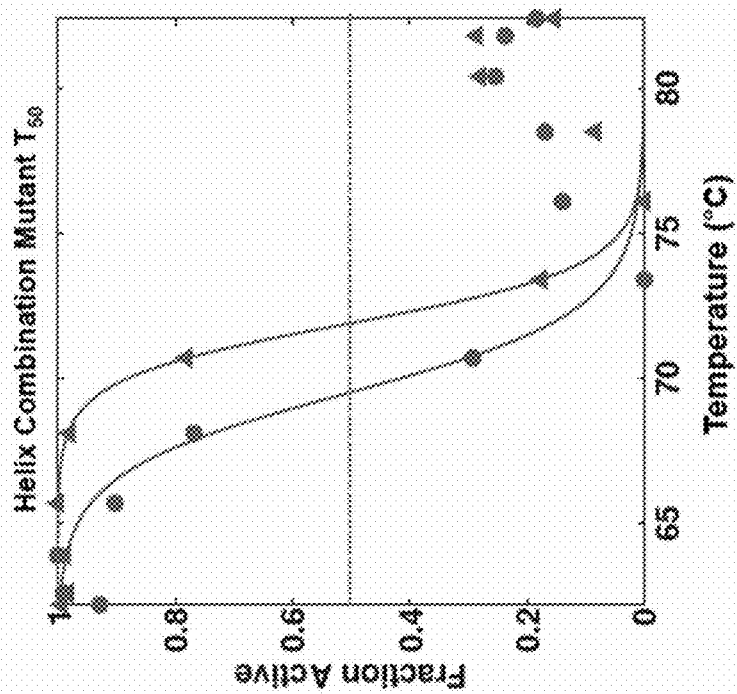


FIG. 10A

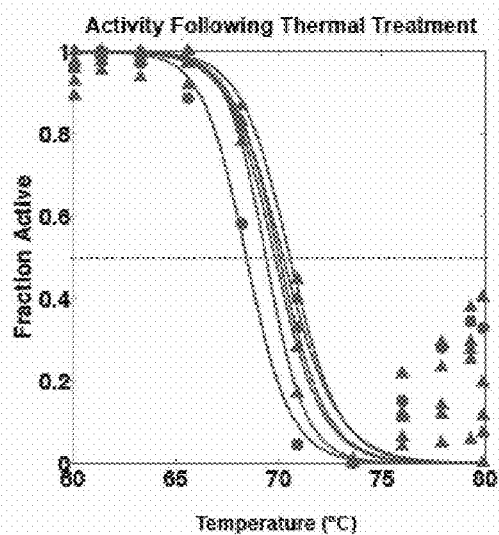


FIG. 10B

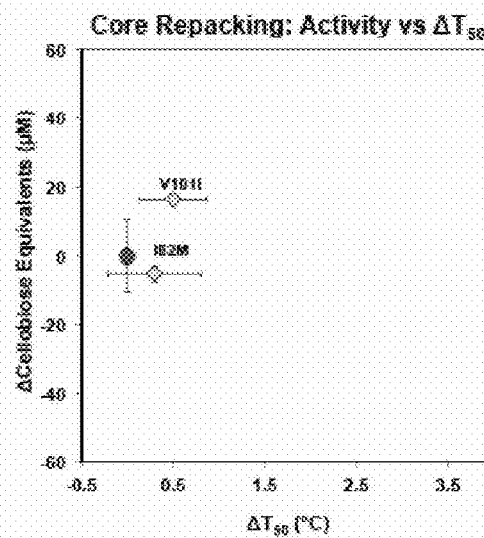


FIG. 10C

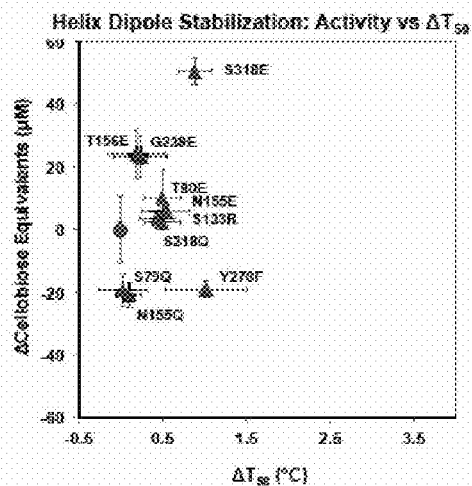


FIG. 11A

FoldX Screen

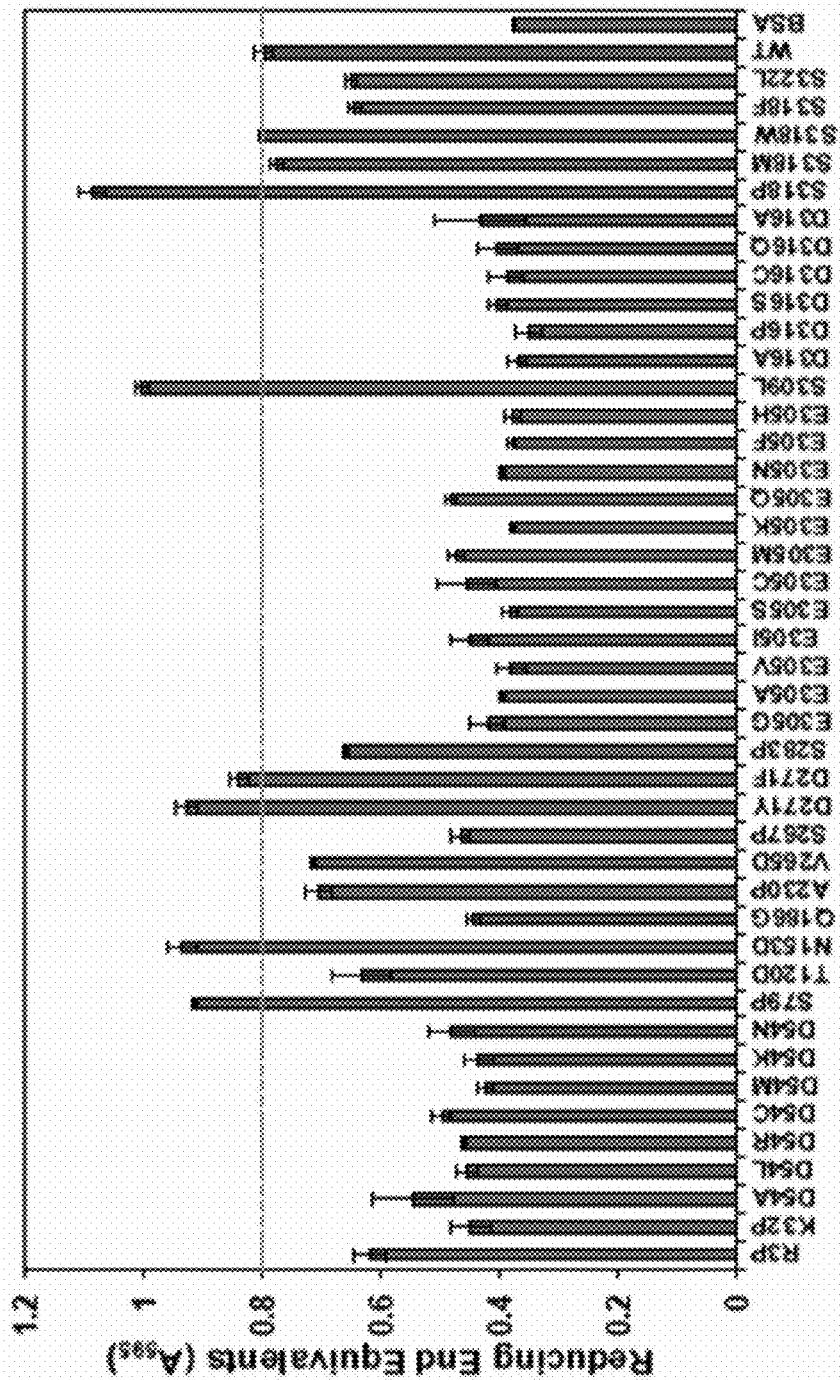


FIG. 11B

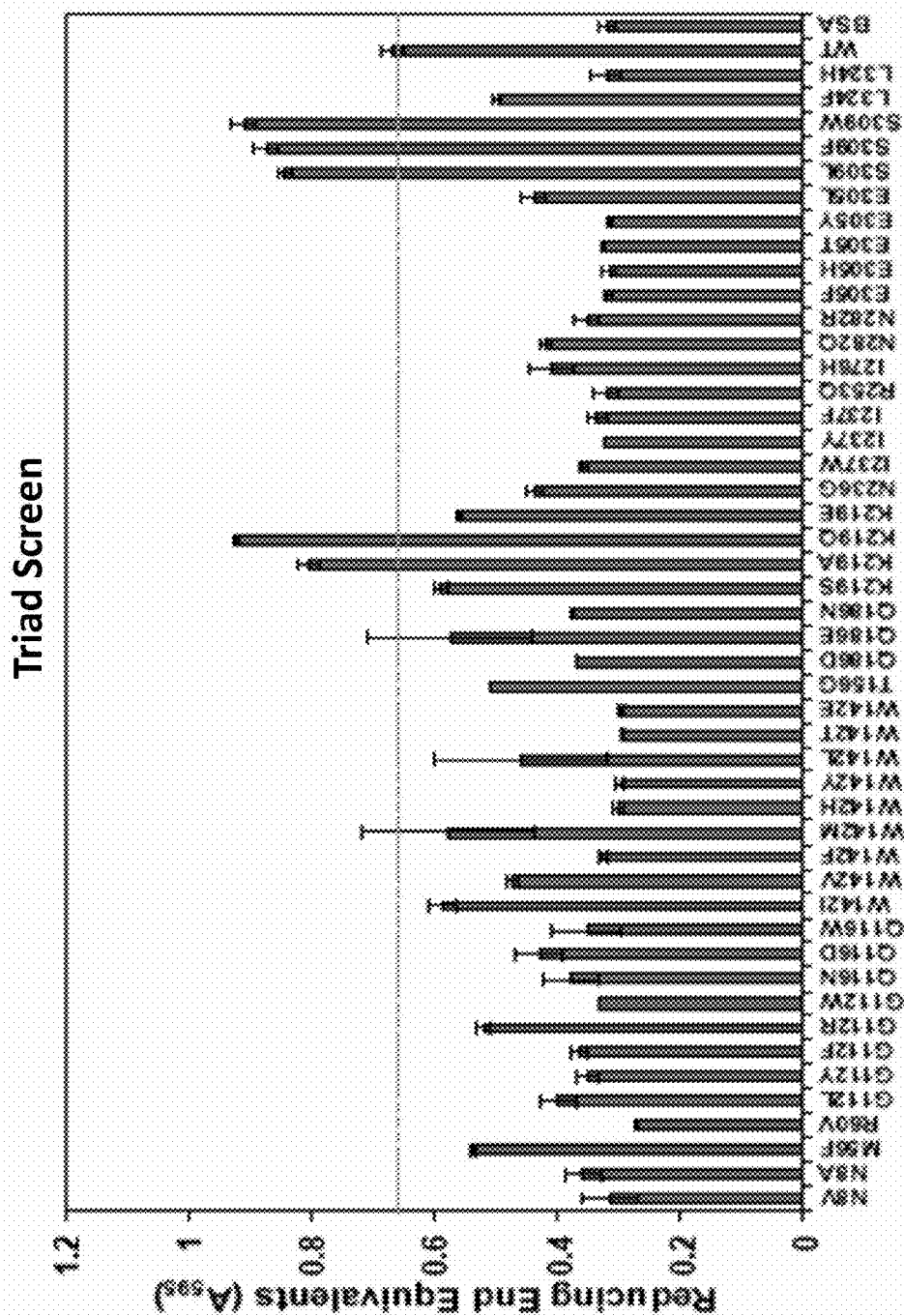


FIG. 12A

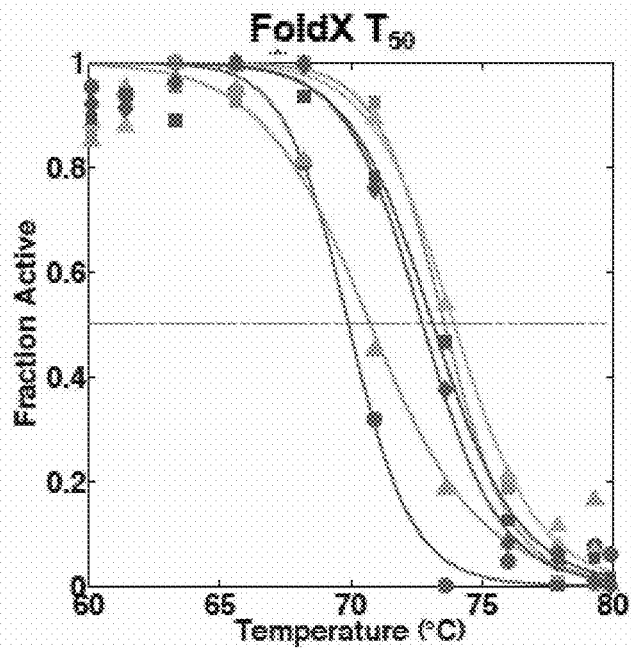


FIG. 12B

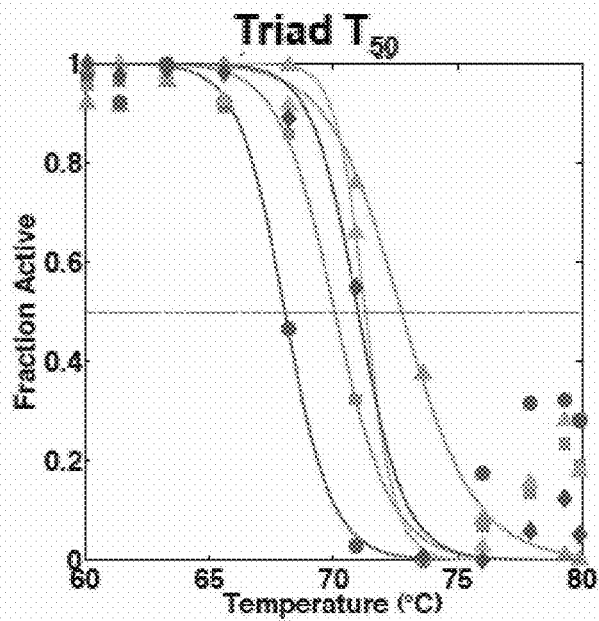




FIG. 13A

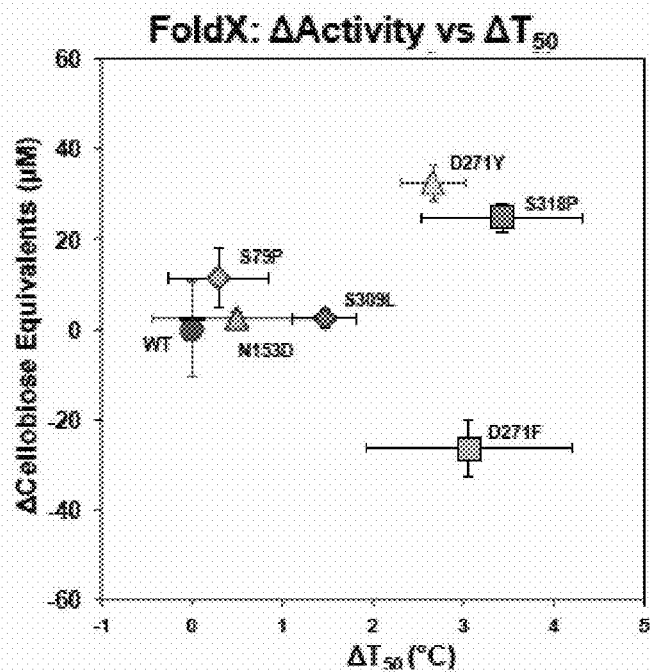


FIG. 13B

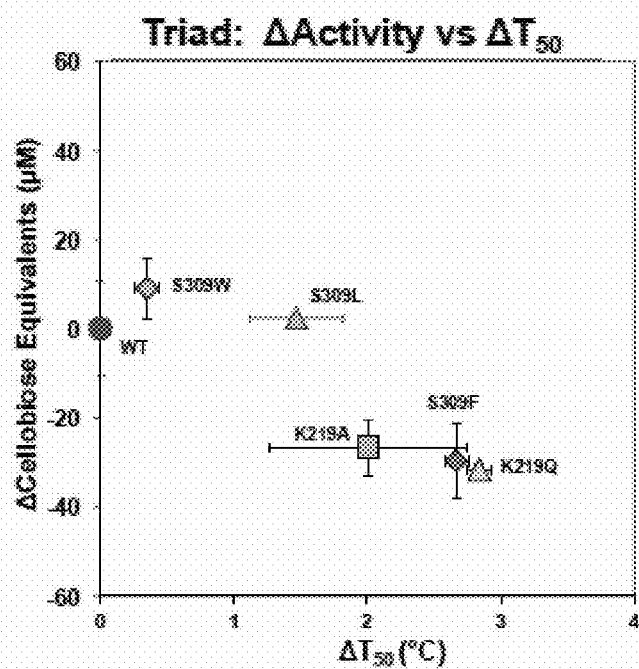


FIG. 14A

## Gly to Xaa Screen

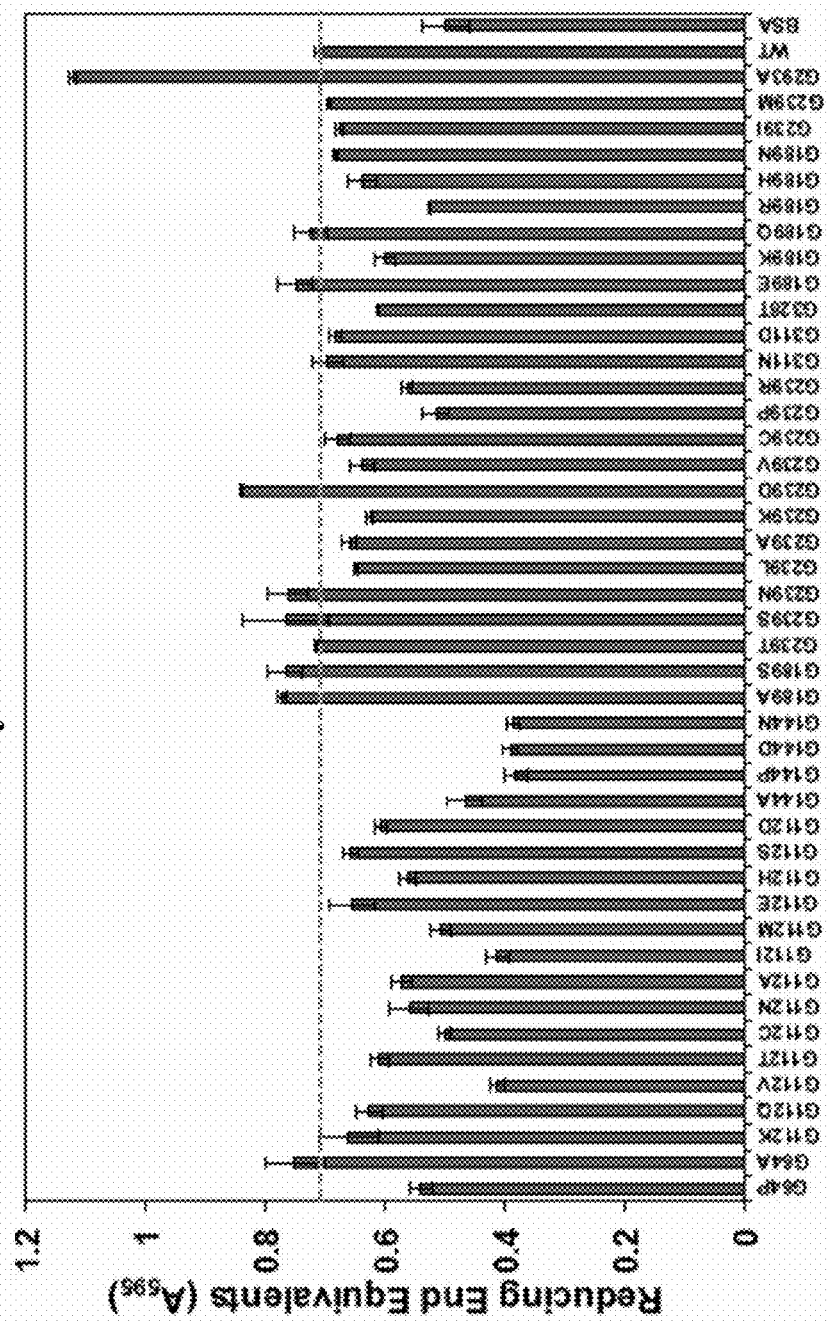


FIG. 14B

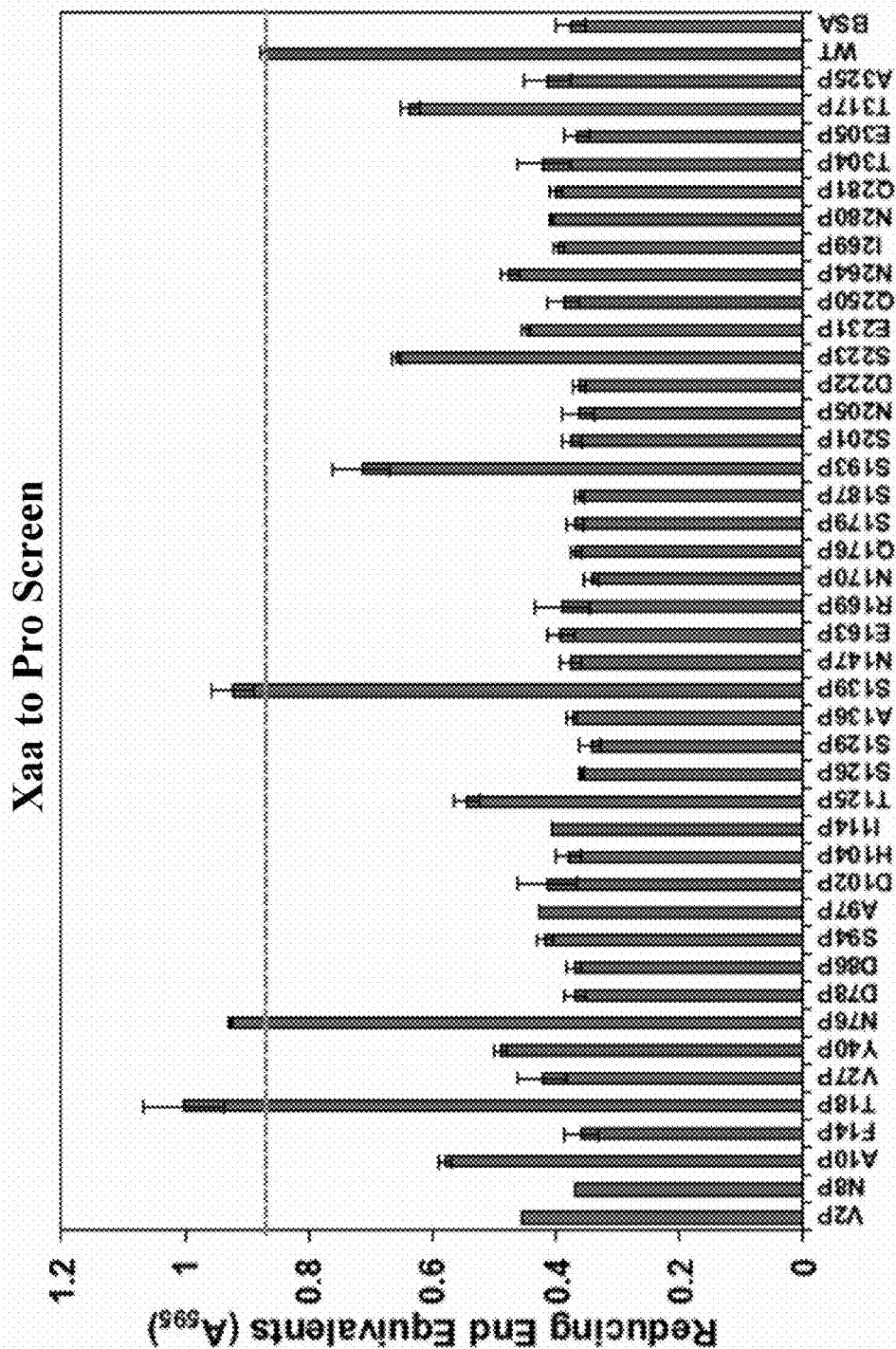


FIG. 15B

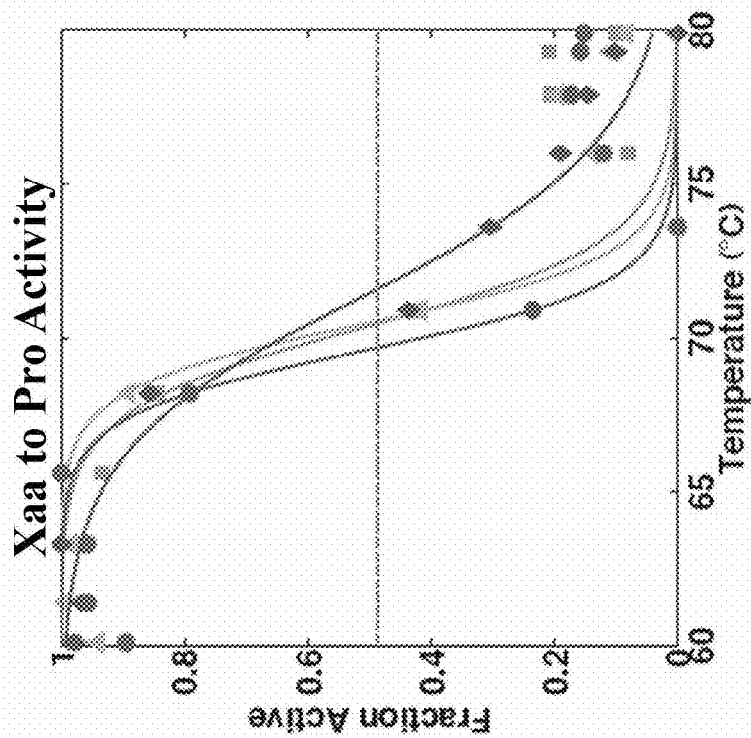


FIG. 15A

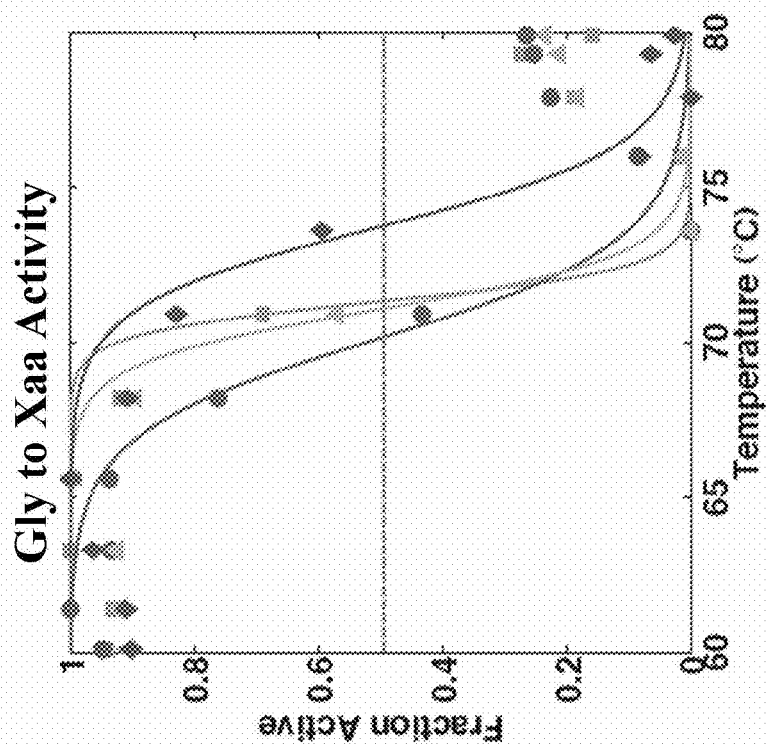


FIG. 16A

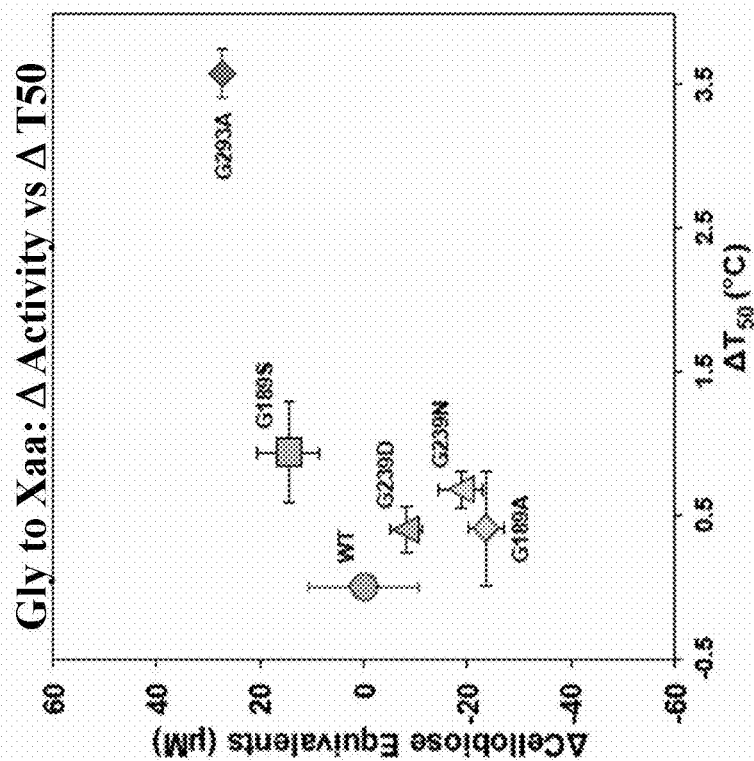


FIG. 16B

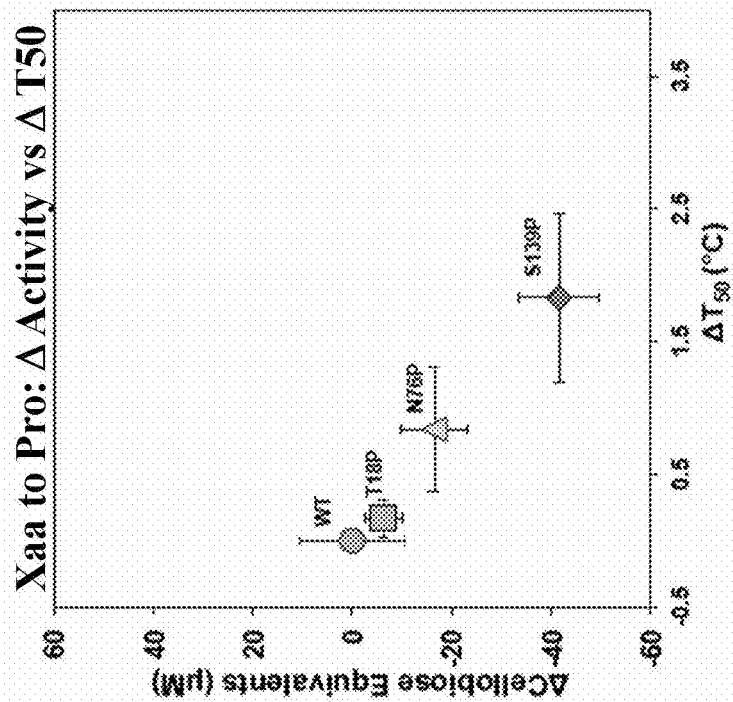


FIG. 17

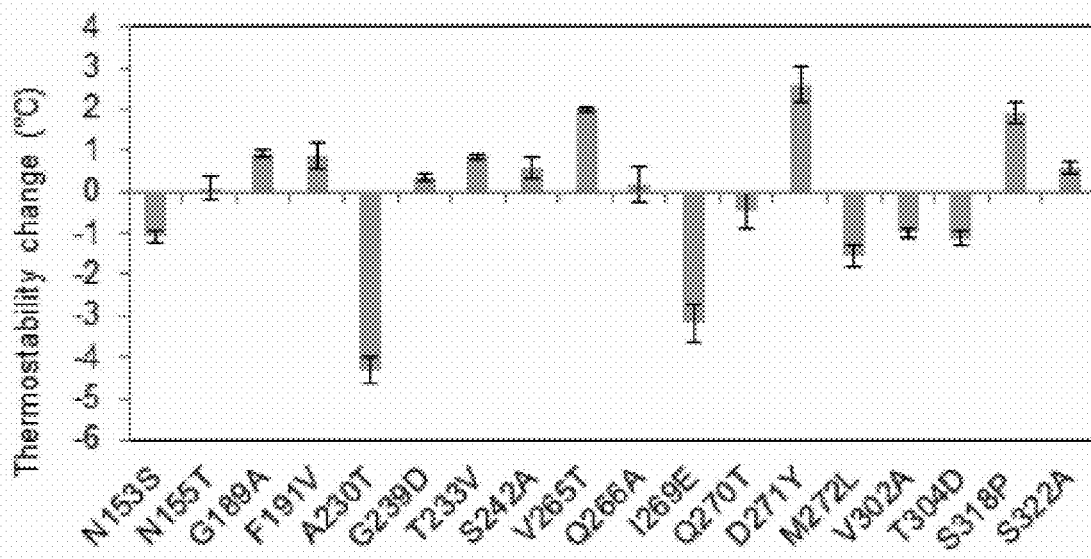


FIG. 18A

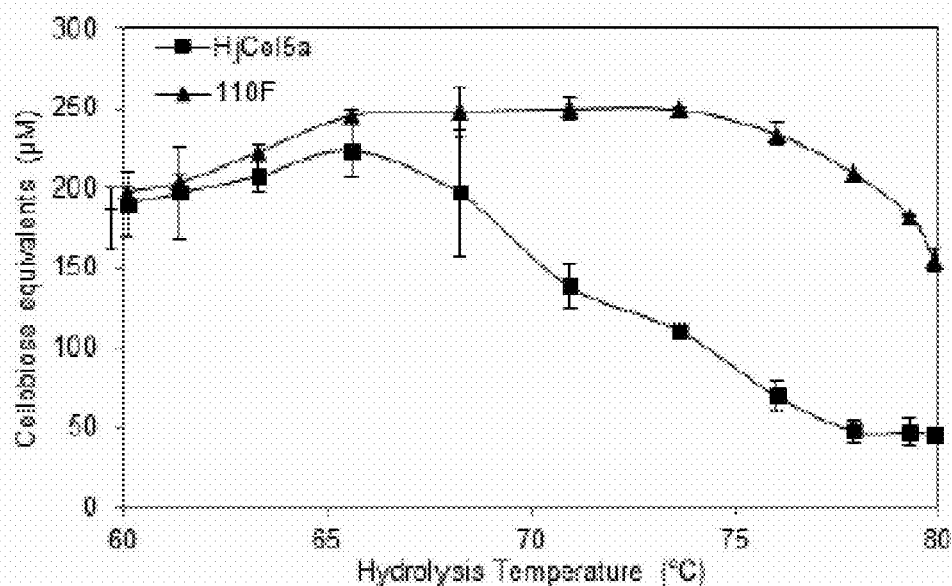


FIG. 18B

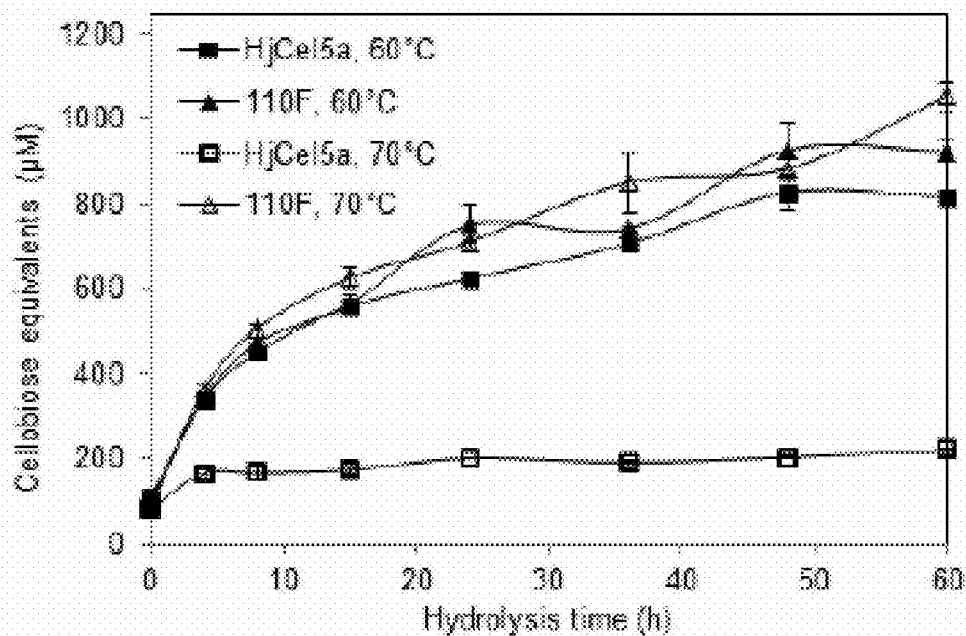


FIG. 19B

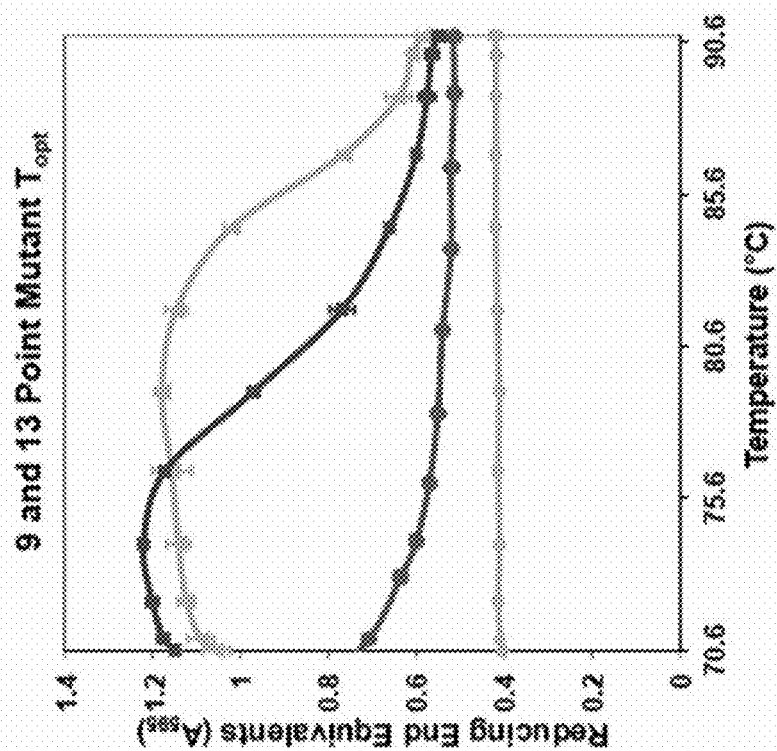


FIG. 19A

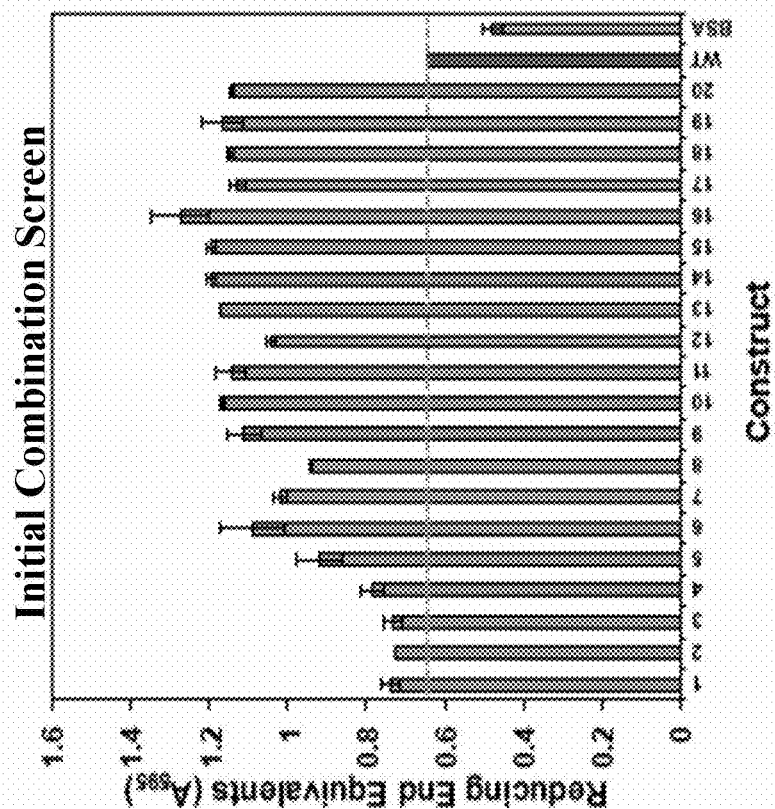




FIG. 19D

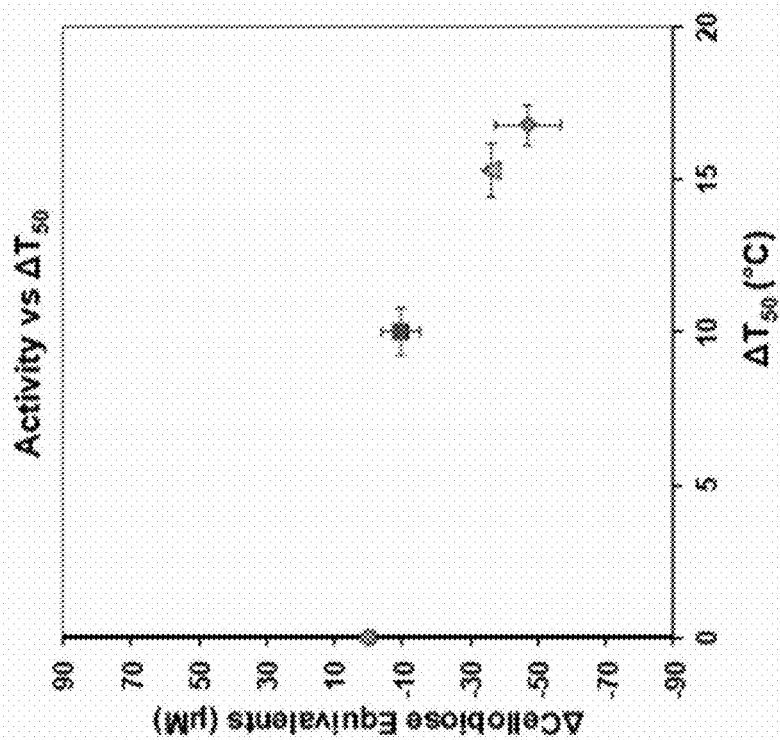


FIG. 19C

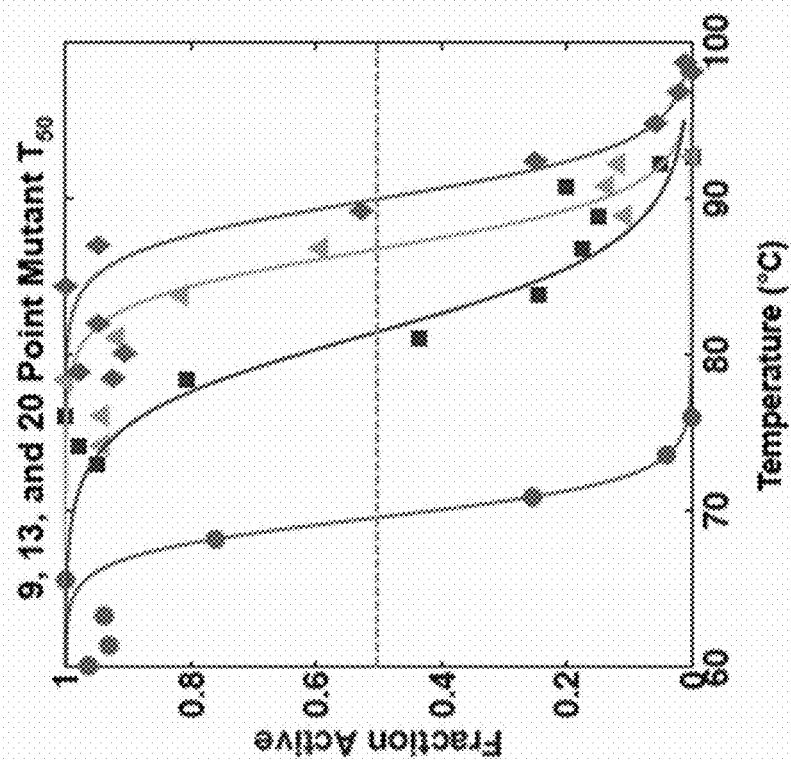


FIG. 19F

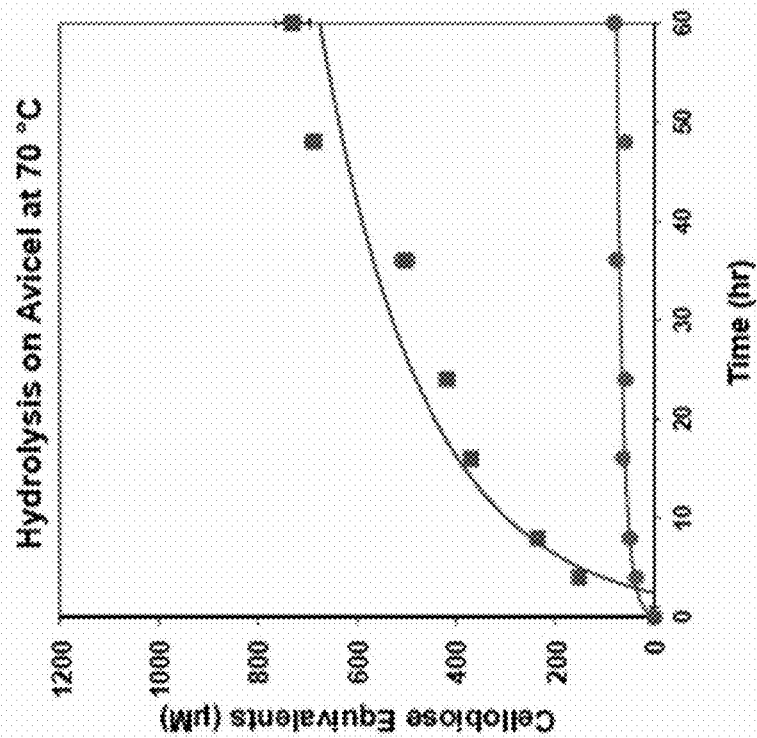


FIG. 19E

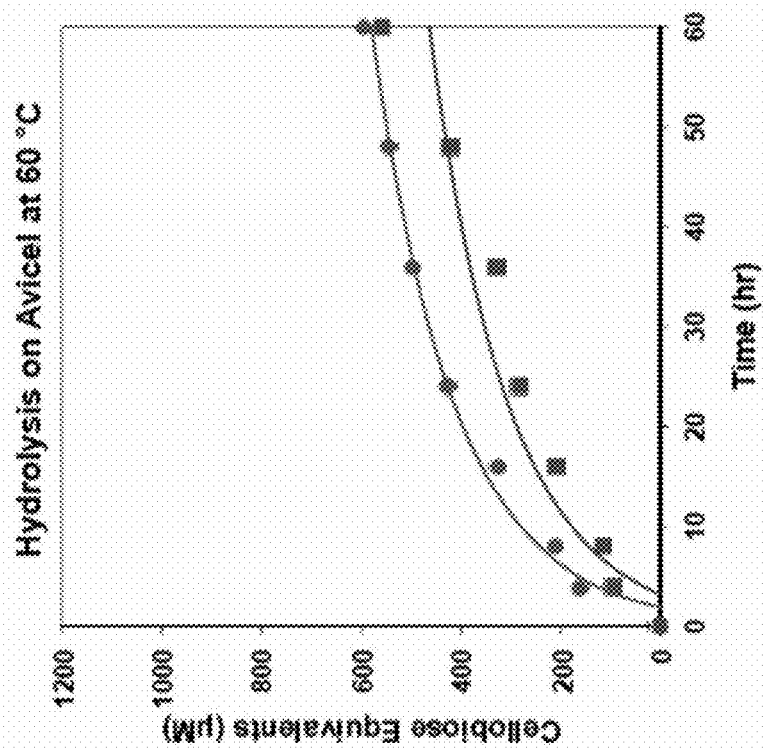


FIG. 20B

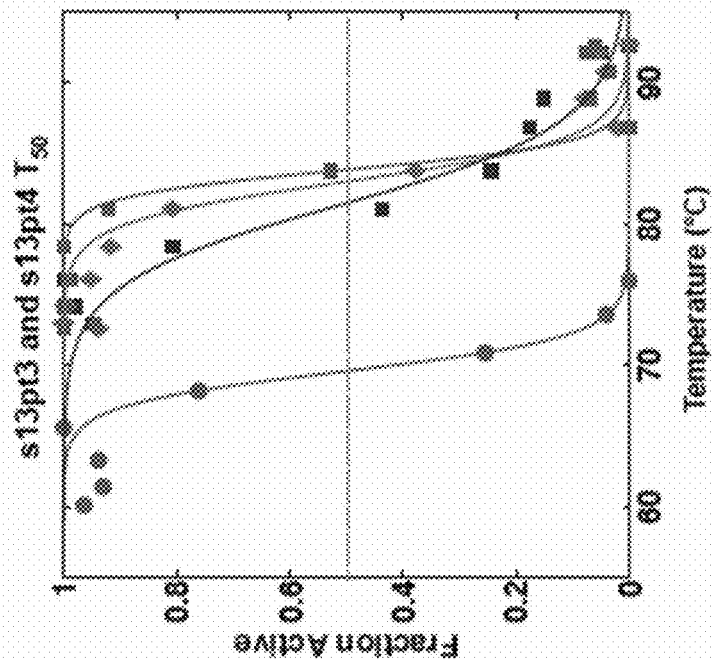


FIG. 20A

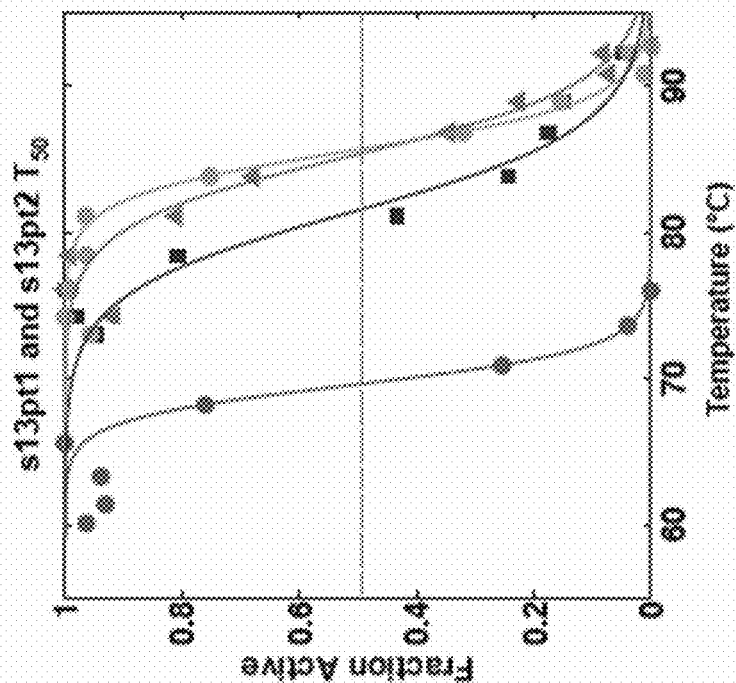


FIG. 20D

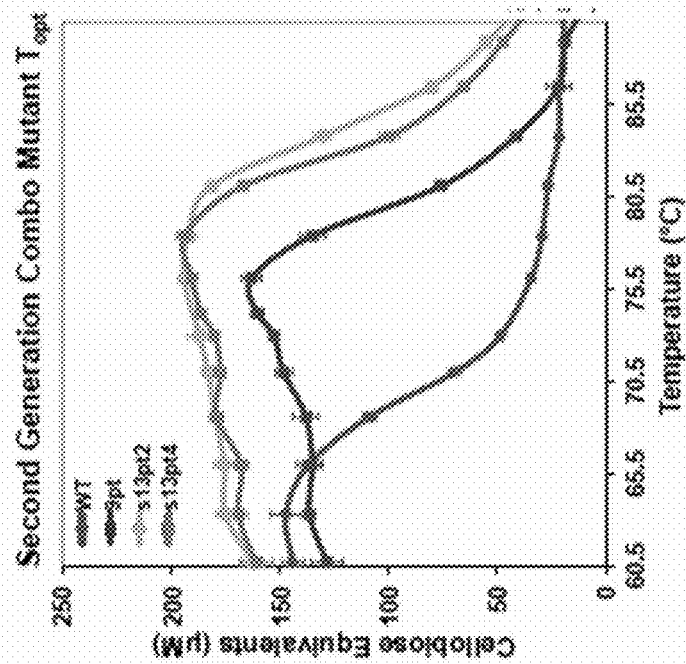


FIG. 20C

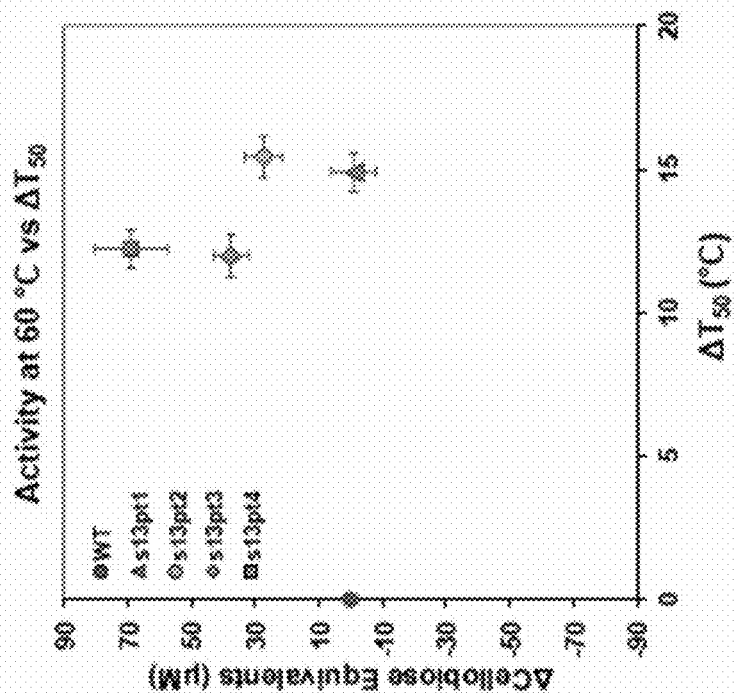


FIG. 21B

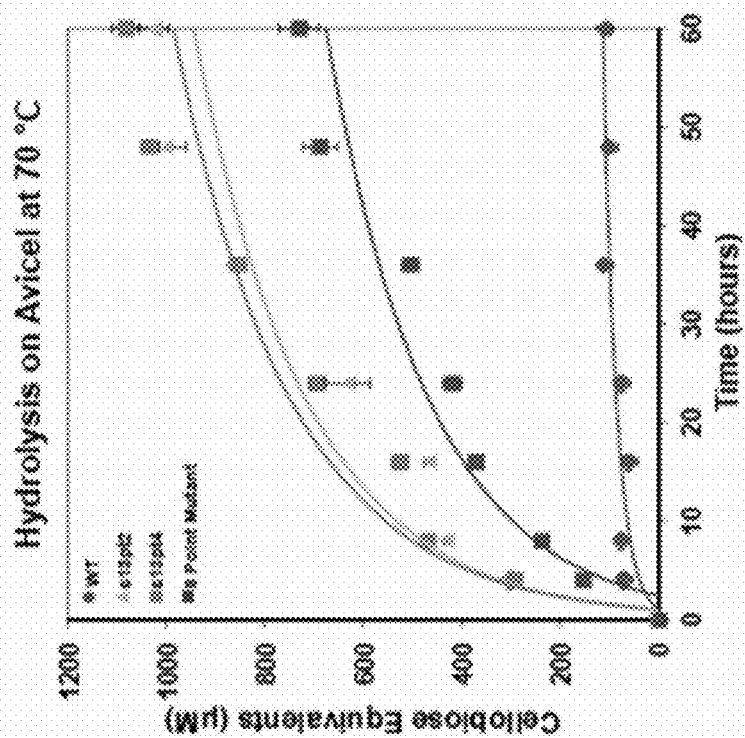


FIG. 21A

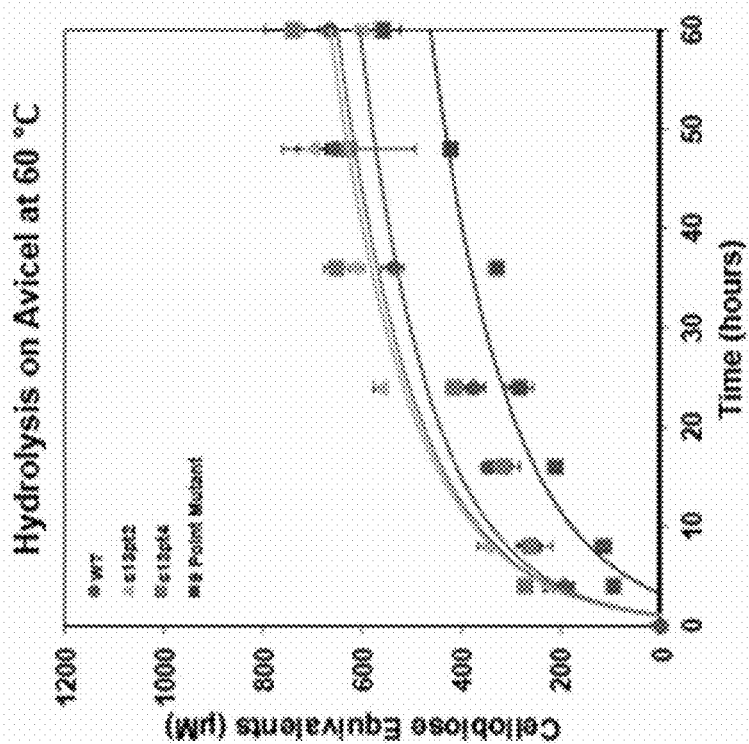


FIG. 21D

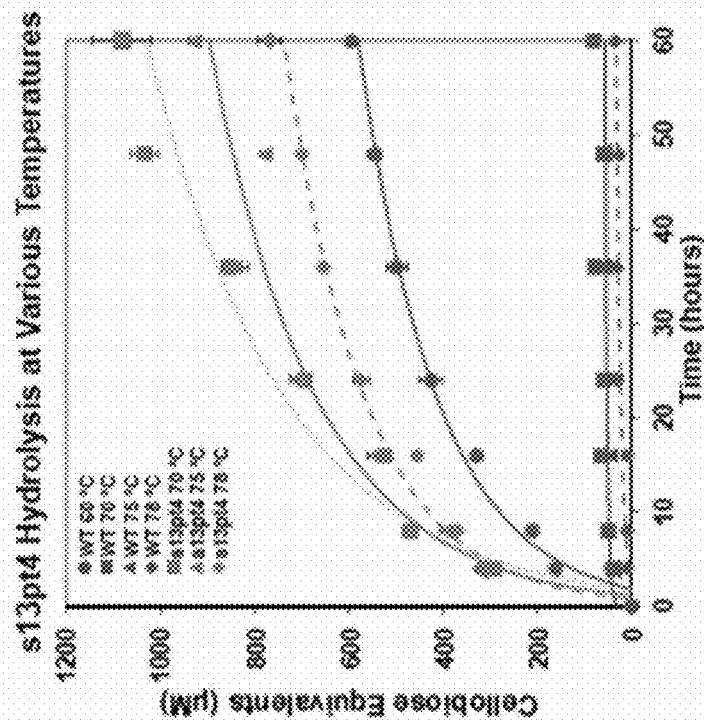


FIG. 21C

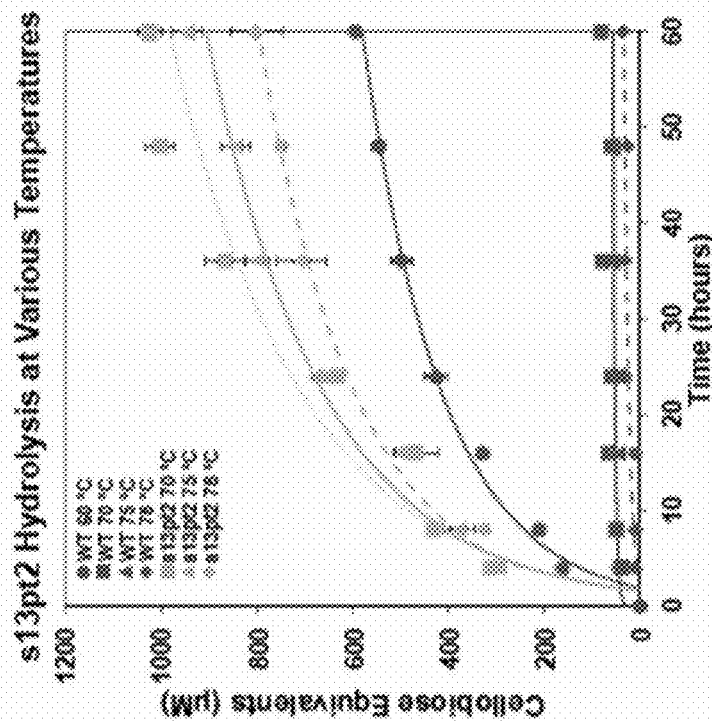


FIG. 22A

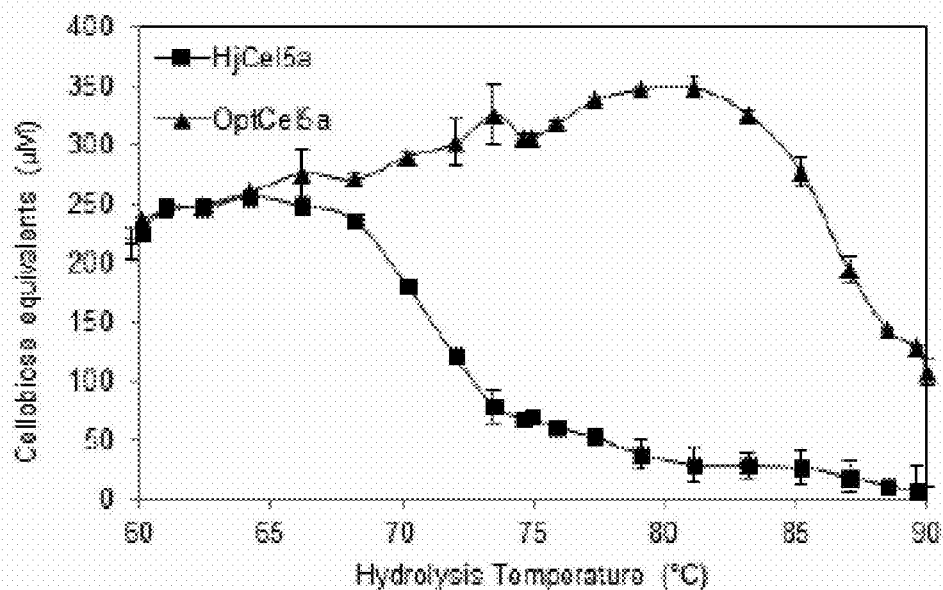
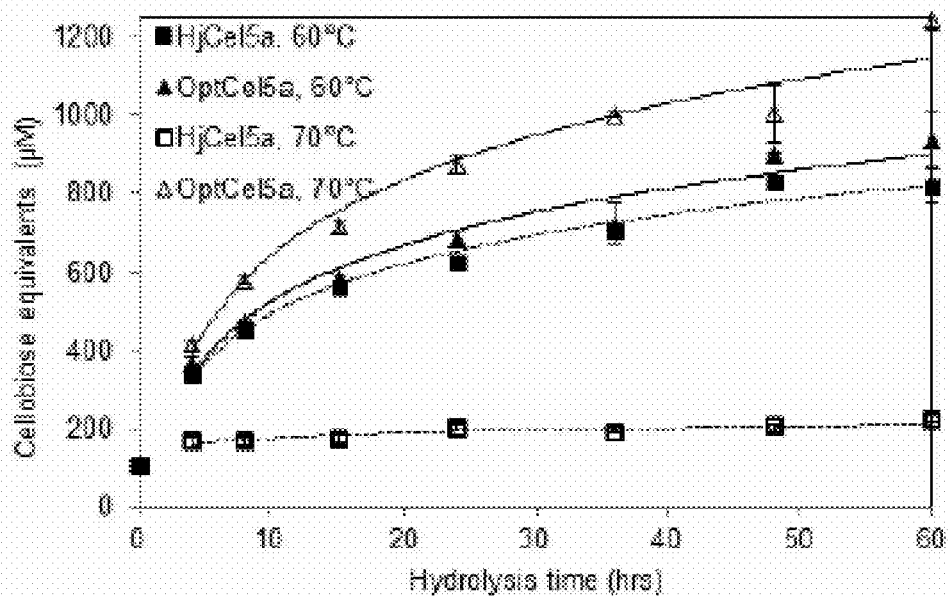


FIG. 22B



1

**ENDOGLUCANASE HAVING ENHANCED  
THERMOSTABILITY AND ACTIVITY****CROSS-REFERENCE TO RELATED  
APPLICATION(S)**

The present application claims priority to and the benefit of U.S. Provisional Application Ser. No. 61/809,144 filed on Apr. 5, 2013 and U.S. Provisional Application Ser. No. 61/822,489 filed on May 13, 2013, the entire contents of both of which are incorporated herein by reference.

**SEQUENCE LISTING**

The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, was amended on Dec. 21, 2015, is named SEQLIST14247221.txt and is 11,300 bytes in size.

**BACKGROUND**

The primary component of plant cell walls, cellulose, is not only the most abundant biopolymer, but it is also a completely renewable resource. In order to expose cellulose, the ligno-cellulosic material made up of cellulose, hemicellulose and lignin, often requires a harsh chemical or temperature-based pretreatment to remove the hemicellulose and lignin. The cellulose is subsequently digested with mixtures of expensive cellulose-digesting enzymes, called cellulases. Digestion (by hydrolysis) of the cellulose by cellulases produces glucose which can be fermented into liquid fuel.

At this time, for the digestion of cellulose to be economically feasible, the cellulase production costs must decrease. For complete cellulose digestion, a mixture of wild type fungal cellulases is known to require at least three different activities: endoglucanase, cellobiohydrolase I, and cellobiohydrolase II, where each type of activity may be carried out by more than one enzyme. The filamentous fungus, *Hypocrea jecorina* (*Trichoderma reesei*) digests cellulose using the least number of cellulase enzymes, and the majority of the digestion is attributed to four enzymes: Cel7A, Cel6A, Cel5A, and Cel7B. As such, an enzymatic mixture made using these *H. jecorina* cellulases requires fewer components for complete digestion.

**SUMMARY**

In some embodiments of the present invention, a composition includes a variant Cel5a endoglucanase having a temperature at which half of the maximal protein activity remains ( $T_{50}$ ) above a  $T_{50}$  of wild type Cel5a endoglucanase and increased enzymatic activity relative to wild type Cel5a endoglucanase.

In some embodiments of the present invention, the variant Cel5a endoglucanase having a  $T_{50}$  above a  $T_{50}$  of wild type Cel5a endoglucanase and increased enzymatic activity includes a single point mutation in the Cel5a amino acid sequence selected from the group consisting of E53D, T57N, S79P, T80E, V101I, S133R, N153D, N155E, T156E, G189S, F191V, T233V, G239E, V265T, D271Y, G293A, S309W/L, S318P/Q/E, and combinations thereof, with mutant residue numbering based on SEQ ID NO: 5.

In some embodiments of the present invention, the variant Cel5a endoglucanase having a  $T_{50}$  above a  $T_{50}$  of wild type Cel5a endoglucanase and increased enzymatic activity includes a combination of point mutations in the Cel5a amino

2

acid sequence. In some embodiments, for example, the combination of point mutations may include any of the following combinations, with mutant residue numbering based on SEQ ID NO: 5:

5 G189A, F191V, G239D, S242A, V265T, D271Y, S318P/E, and S322A;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309L, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309W, and S318P;

10 E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309L, and S318E;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309W, and S318E; or

15 E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, F191V, T233V, G239E, V265T, D271Y, G293A, S309W, and S318P.

In some embodiments of the present invention, a method of hydrolyzing cellulose includes hydrolyzing the cellulose with the variant Cel5a endoglucanase composition having a  $T_{50}$  above a  $T_{50}$  of wild type Cel5a endoglucanase and increased enzymatic activity relative to wild type.

In some embodiments of the present invention, a composition includes a variant Cel5a endoglucanase having a  $T_{50}$  above the  $T_{50}$  of wild type Cel5a endoglucanase, the composition including a point mutation selected from D13E, T18P, E53D, T57N, N76P, S79P/Q, T80E, I82M, V101I, S133R, S139P, N153D, N155E/Q, T156E, G189S/A/E, F191V, K219Q/A, T233V, G239E/N/D, V265T, D271Y/F, Y278F, G293A, S309W/F/L, S318P/Q/E, or any combination thereof, with mutant residue numbering based on SEQ ID NO: 5.

In some embodiments of the present invention, a composition includes a variant Cel5a endoglucanase having increased protein expression in a host organism compared to wild type Cel5a, the composition including a point mutation selected from D13E, T18P, G64A, N76P, S79P/Q/E, T80E/Q, I82L/M, V101I/L, A122E, S133R, S139P, N155E/Q, T156E, G189S/A/E/K, K219Q/A, G239E/N/D, V265T, D271Y/F, Y278F, G293A, S309W/F/L, or S318P/Q/E, with mutant residue numbering based on SEQ ID NO: 5.

In some embodiments of the present invention, a composition includes a variant Cel5a endoglucanase having a  $T_{50}$  above the  $T_{50}$  of wild type Cel5a endoglucanase, increased enzymatic activity relative to wild type Cel5a endoglucanase, and increased protein expression in a host organism relative to wild type Cel5a, the variant Cel5a endoglucanase having a combination of mutations which may include any of the following combinations, with mutant residue numbering based on SEQ ID NO: 5:

50 G189A, F191V, G239D, S242A, V265T, D271Y, S318P/E, and S322A;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309L, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309W, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309L, and S318E;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309W, and S318E; or

60 E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, F191V, T233V, G239E, V265T, D271Y, G293A, S309W, and S318P.

**BRIEF DESCRIPTION OF THE DRAWINGS**

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application



publication with color drawings will be provided by the Office upon request and payment of the necessary fee.

These and other features and advantages of the present invention will be better understood by reference to the following detailed description when considered in conjunction with the accompanying drawings.

FIG. 1 is a graph measuring thermostability and activity in wild type (WT) and the indicated Cel5a point mutant proteins (with mutant residue numbering based on SEQ ID NO: 5) following a reducing end (Park-Johnson) assay as described herein, with bovine serum albumin (BSA) used as a control, according to embodiments of the present invention.

FIG. 2 is a graph corresponding to thermostability and activity of FIG. 1 showing the amount of activity versus an increase in temperature of the wild type (green) and the indicated Cel5a mutants (with mutant residue numbering based on SEQ ID NO: 5): D13E (pink circles), E53D (dark blue diamonds), T57N (light blue triangles), G189A (yellow triangles), and G293A (orange squares), where the dashed line indicates the point at which 50% of the maximal protein activity remains ( $T_{50}$ ), according to embodiments of the present invention.

FIG. 3A is a  $T_{50}$  plot for the Cel5a D13E mutant of FIG. 2, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (blue) are shown, according to embodiments of the present invention.

FIG. 3B is a  $T_{50}$  plot for the Cel5a E53D mutant of FIG. 2, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (blue) are shown, according to embodiments of the present invention.

FIG. 3C is a  $T_{50}$  plot for the Cel5a T57N mutant of FIG. 2, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (blue) are shown, according to embodiments of the present invention.

FIG. 3D is a  $T_{50}$  plot for the Cel5a G189A mutant of FIG. 2, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (blue) are shown, according to embodiments of the present invention.

FIG. 3E is a  $T_{50}$  plot for the Cel5a G293A mutant of FIG. 2, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (blue) are shown, according to embodiments of the present invention.

FIG. 3F is a  $T_{50}$  plot for the Cel5a I82L mutant of FIG. 2, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (blue) are shown, according to embodiments of the present invention.

FIG. 3G is a  $T_{50}$  plot for the Cel5a V101L mutant of FIG. 2, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (blue) are shown, according to embodiments of the present invention.

FIG. 4 is a graph showing the cellulase activity of the indicated Cel5a point mutants (with mutant residue numbering based on SEQ ID NO: 5) versus the  $\Delta T_{50}$  (compared to wild type Cel5a), where WT (green circle), D13E (pink circle), E53D (dark blue diamond), T57N (light blue triangle), G189A (yellow triangle), and G293A (orange square) are shown, according to embodiments of the present invention.

FIG. 5 is a graph measuring thermostability and activity in wild type (WT) and the indicated core repacking Cel5a point

mutant proteins (with mutant residue numbering based on SEQ ID NO: 5) following a reducing end (Park-Johnson) assay as described herein, with bovine serum albumin (BSA) used as a control, according to embodiments of the present invention.

FIG. 6A is a  $T_{50}$  plot for the Cel5a I82L mutant of FIG. 5, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 6B is a  $T_{50}$  plot for the Cel5a V101I mutant of FIG. 5, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 7 is a graph measuring thermostability and activity in wild type (WT) and the indicated helix dipole Cel5a point mutant proteins (with mutant residue numbering based on SEQ ID NO: 5) following a reducing end (Park-Johnson) assay as described herein, with bovine serum albumin (BSA) used as a control, according to embodiments of the present invention.

FIG. 8A is a  $T_{50}$  plot for the Cel5a S79Q mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8B is a  $T_{50}$  plot for the Cel5a T80E mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8C is a  $T_{50}$  plot for the Cel5a S133R mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8D is a  $T_{50}$  plot for the Cel5a N155E mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8E is a  $T_{50}$  plot for the Cel5a N155Q mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8F is a  $T_{50}$  plot for the Cel5a T156E mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8G is a  $T_{50}$  plot for the Cel5a G239E mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

5

FIG. 8H is a  $T_{50}$  plot for the Cel5a Y278F mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8I is a  $T_{50}$  plot for the Cel5a S318E mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8J is a  $T_{50}$  plot for the Cel5a S318Q mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8K is a  $T_{50}$  plot for the Cel5a S79E mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8L is a  $T_{50}$  plot for the Cel5a T80Q mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8M is a  $T_{50}$  plot for the Cel5a A122E mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 8N is a  $T_{50}$  plot for the Cel5a G239Q mutant of FIG. 7, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green) and the point mutant (yellow) are shown, according to embodiments of the present invention.

FIG. 9A is a graph measuring the reducing ends equivalents (Park-Johnson assay) for Cel5a WT (green circles) and Helix combo mutant (blue triangles) proteins (with mutant residue numbering based on SEQ ID NO: 5) over a range of temperatures as indicated, according to embodiments of the present invention.

FIG. 9B is a  $T_{50}$  plot for the Cel5a WT and Helix combo mutant of FIG. 9A, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green circles) and the Helix combo mutant (blue triangles) are shown, according to embodiments of the present invention.

FIG. 10A is a  $T_{50}$  plot for the Cel5a WT and core repacking point mutations (with mutant residue numbering based on SEQ ID NO: 5) having a  $\Delta T_{50} > 0.5^\circ$  C., where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (green circles) and core repacking mutants (blue triangles) are shown, according to embodiments of the present invention.

FIG. 10B is a graph showing the cellulase activity of the indicated core repacking Cel5a point mutants (with mutant residue numbering based on SEQ ID NO: 5) versus the  $\Delta T_{50}$  (compared to wild type Cel5a), where WT (green circle) and

6

the indicated core repacking mutants (yellow diamonds) are shown, according to embodiments of the present invention.

FIG. 10C is a graph showing the cellulase activity of the indicated helix dipole Cel5a point mutants (with mutant residue numbering based on SEQ ID NO: 5) versus the  $\Delta T_{50}$  (compared to wild type Cel5a), where WT (green circle) and the indicated helix dipole mutants (blue diamonds) are shown, according to embodiments of the present invention.

FIG. 11A is a graph measuring thermostability and activity in wild type (WT) and the indicated Cel5a point mutant proteins (with mutant residue numbering based on SEQ ID NO: 5) selected from a FoldX screen as described herein, with bovine serum albumin (BSA) used as a control, according to embodiments of the present invention.

FIG. 11B is a graph measuring thermostability and activity in wild type (WT) and the indicated Cel5a point mutant proteins (with mutant residue numbering based on SEQ ID NO: 5) selected from a Triad screen as described herein, with bovine serum albumin (BSA) used as a control, according to embodiments of the present invention.

FIG. 12A is a graph corresponding to thermostability and activity of FIG. 11A showing the amount of activity versus an increase in temperature of the wild type (green) and the indicated FoldX Cel5a mutants (with mutant residue numbering based on SEQ ID NO: 5): N153D (pink triangles), D271F (orange squares), D271Y (yellow triangles), S309L (blue diamonds), and S319P (green squares), where the dashed line indicates the point at which 50% of the maximal protein activity persists ( $T_{50}$ ), according to embodiments of the present invention.

FIG. 12B is a graph corresponding to thermostability and activity of FIG. 11B showing the amount of activity versus an increase in temperature of the wild type (green) and the indicated Triad Cel5a mutants (with mutant residue numbering based on SEQ ID NO: 5): K219A (orange squares), K219Q (yellow triangles), S309F (blue diamonds), and S309L (pink triangles), where the dashed line indicates the point at which 50% of the maximal protein activity persists ( $T_{50}$ ), according to embodiments of the present invention.

FIG. 13A is a graph showing the cellulase activity of the indicated FoldX Cel5a point mutants (with mutant residue numbering based on SEQ ID NO: 5) versus the  $\Delta T_{50}$  (compared to wild type Cel5a), as indicated in FIG. 12A, according to embodiments of the present invention.

FIG. 13B is a graph showing the cellulase activity of the indicated Triad Cel5a point mutants (with mutant residue numbering based on SEQ ID NO: 5) versus the  $\Delta T_{50}$  (compared to wild type Cel5a), as indicated in FIG. 12B, according to embodiments of the present invention.

FIG. 14A is a graph measuring thermostability and activity in wild type (WT) and the indicated Glycine to Xaa Cel5a point mutant proteins (with mutant residue numbering based on SEQ ID NO: 5) following a reducing end (Park-Johnson) assay as described herein, with bovine serum albumin (BSA) used as a control, according to embodiments of the present invention.

FIG. 14B is a graph measuring thermostability and activity in wild type (WT) and the indicated Xaa to Proline Cel5a point mutant proteins (with mutant residue numbering based on SEQ ID NO: 5) following a reducing end (Park-Johnson) assay as described herein, with bovine serum albumin (BSA) used as a control, according to embodiments of the present invention.

FIG. 15A is a  $T_{50}$  plot for the Cel5a WT and Gly to Xaa point mutations (with mutant residue numbering based on SEQ ID NO: 5), where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treat-

7

ment from 60-80° C., where WT (grey circles) and G189S (orange squares), G239N (yellow triangles), and G293A (blue diamonds) are shown, according to embodiments of the present invention.

FIG. 15B is a  $T_{50}$  plot for the Cel5a WT and Xaa to Proline point mutations (with mutant residue numbering based on SEQ ID NO: 5), where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-80° C., where WT (grey circles) and T18P (orange squares), N76P (yellow triangles), and S139P (blue diamonds) are shown, according to embodiments of the present invention.

FIG. 16A is a graph showing the  $\Delta$  cellulase activity (compared to wild type Cel5a) of the indicated Gly to Xaa Cel5a point mutants (with mutant residue numbering based on SEQ ID NO: 5) versus the  $\Delta T_{50}$  (compared to wild type Cel5a), as indicated in FIG. 15A, according to embodiments of the present invention.

FIG. 16B is a graph showing the  $\Delta$  cellulase activity (compared to wild type Cel5a) of the indicated Xaa to Proline Cel5a point mutants (with mutant residue numbering based on SEQ ID NO: 5) versus the  $\Delta T_{50}$  (compared to wild type Cel5a), as indicated in FIG. 15B, according to embodiments of the present invention.

FIG. 17 is a graph showing the  $\Delta$  thermostability (compared to wild type Cel5a) of the indicated Cel5a point mutants (with mutant residue numbering based on SEQ ID NO: 5), according to embodiments of the present invention.

FIG. 18A is a graph showing the amount of cellulase activity versus hydrolysis temperature for WT (HjCel5a) (squares) and 110F Cel5a combination mutant (triangles) (with mutant residue numbering based on SEQ ID NO: 5), as described herein, according to embodiments of the present invention.

FIG. 18B is a graph showing the amount of cellulase activity over 60 hours for WT (HjCel5a) (squares) and 110F Cel5a combination mutant (triangles) (with mutant residue numbering based on SEQ ID NO: 5) at 60° C. and 70° C. as indicated, according to embodiments of the present invention.

FIG. 19A is a graph of thermostability and activity in wild type (WT) and the indicated Cel5a combination mutant proteins (with mutant residue numbering based on SEQ ID NO: 5) following a reducing end (Park-Johnson) assay as described herein, with bovine serum albumin (BSA) used as a control, according to embodiments of the present invention.

FIG. 19B is a graph of the reducing ends equivalents (Park-Johnson assay) for Cel5a WT (green circles), 9-point mutant (blue squares), and 13-point mutant (yellow squares) proteins (with mutant residue numbering based on SEQ ID NO: 5) over a range of temperatures as indicated, according to embodiments of the present invention.

FIG. 19C is a  $T_{50}$  plot for the 9-, 13-, and 20-point mutants of FIG. 19A, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-100° C., where WT (green circles), 9-point mutant (blue squares), 13-point mutant (yellow squares), and 20-point mutant (purple diamonds) are shown, according to embodiments of the present invention.

FIG. 19D is a graph showing the  $\Delta$  cellulase activity (compared to wild type Cel5a) of the indicated 9-point mutant (with mutant residue numbering based on SEQ ID NO: 5) versus the  $\Delta T_{50}$  (compared to wild type Cel5a) at 60° C. from a single point assay as described herein, according to embodiments of the present invention.

FIG. 19E is a graph showing the amount of cellulase activity of WT (green circles) and the indicated 9-point mutant

8

(blue circles) (with mutant residue numbering based on SEQ ID NO: 5) at 60° C., according to embodiments of the present invention.

FIG. 19F is a graph showing the amount of cellulase activity of WT (green circles) and the indicated 9-point mutant (blue circles) (with mutant residue numbering based on SEQ ID NO: 5) at 70° C., according to embodiments of the present invention.

FIG. 20A is a  $T_{50}$  plot for the s13pt1 and s13pt2 combination mutants, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-100° C., where WT (green circles), 9-point mutant (blue squares), s18pt1 mutant (red triangles), and s13pt2 mutant (yellow circles) are shown (with mutant residue numbering based on SEQ ID NO: 5), according to embodiments of the present invention.

FIG. 20B is a  $T_{50}$  plot for the s13pt3 and s13pt4 combination mutants, where the fraction active reflects the amount of the total protein that is active after a 10 minute heat treatment from 60-100° C., where WT (green circles), 9-point mutant (blue squares), s18pt3 mutant (turquoise diamonds), and s13pt4 mutant (purple squares) are shown (with mutant residue numbering based on SEQ ID NO: 5), according to embodiments of the present invention.

FIG. 20C is a graph showing the  $\Delta$  cellulase activity (compared to wild type Cel5a) of the indicated combination mutant (with mutant residue numbering based on SEQ ID NO: 5) versus the  $\Delta T_{50}$  (compared to wild type Cel5a) at 60° C. from FIGS. 20A, 20B, according to embodiments of the present invention.

FIG. 20D is a graph showing the amount of cellulase activity versus hydrolysis temperature for WT (green circles), s13pt2 mutant (yellow circles), and s13pt4 mutant (purple squares) (with mutant residue numbering based on SEQ ID NO: 5) as described herein, according to embodiments of the present invention.

FIG. 21A is a graph showing the amount of cellulase activity over 60 hours for WT (green circles), 9-point mutant (blue squares), s13pt2 mutant (yellow circles), and s13pt4 mutant (purple squares) (with mutant residue numbering based on SEQ ID NO: 5) at 60° C., according to embodiments of the present invention.

FIG. 21B is a graph showing the amount of cellulase activity over 60 hours for WT (green circles), 9-point mutant (blue squares), s13pt2 mutant (yellow circles), and s13pt4 mutant (purple squares) (with mutant residue numbering based on SEQ ID NO: 5) at 70° C., according to embodiments of the present invention.

FIG. 21C is a graph showing the amount of cellulase activity over 60 hours for WT (green circles) and s13pt2 mutant (yellow circles) (with mutant residue numbering based on SEQ ID NO: 5) at 60° C. and/or 70° C., 75° C., and 78° C. as indicated, according to embodiments of the present invention.

FIG. 21D is a graph showing the amount of cellulase activity over 60 hours for WT (green circles) and s13pt4 mutant (purple squares) (with mutant residue numbering based on SEQ ID NO: 5) at 60° C. and/or 70° C., 75° C., and 78° C. as indicated, according to embodiments of the present invention.

FIG. 22A is a graph showing the amount of cellulase activity versus hydrolysis temperature for WT (HjCel5a) (squares) and OptCel5a combination mutant (triangles) (with mutant residue numbering based on SEQ ID NO: 5) as described herein, according to embodiments of the present invention.

FIG. 22B is a graph showing the amount of cellulase activity over 60 hours for WT (HjCel5a) (squares) and 110F Cel5a combination mutant (triangles) (with mutant residue number-

ing based on SEQ ID NO: 5) at 60° C. (closed) and 70° C. (open) as indicated, according to embodiments of the present invention.

#### DETAILED DESCRIPTION

Some embodiments of the present invention include variants of the endoglucanase Cel5a having thermostability at higher temperatures compared to wild type (WT) Cel5a. As used herein, “thermostability” refers to the stability and resistance of the enzyme protein (e.g., Cel5a protein) in response to changes in temperature. In this disclosure, an increase in thermostability refers to an increase in the range of higher temperatures at which the enzyme protein is stable and does not degrade compared to wild type. As disclosed herein, a T<sub>50</sub> assay is used to determine the thermostability of a protein. Specifically, a T<sub>50</sub> assay determines the temperature at which half of the maximal protein activity remains. In this way, for example, the Cel5a variants disclosed herein having increased thermostability do not degrade at some of the higher temperatures at which WT Cel5a degrades. For example, an increase in thermostability is any increase in the T<sub>50</sub> of a Cel5a variant protein above that of WT Cel5a protein, resulting in a positive  $\Delta T_{50}$  compared to WT. For example, any increase in a Cel5a variant T<sub>50</sub> may be beneficial. In some embodiments, a Cel5a variant T<sub>50</sub> may be up to about 17° C. higher than wild type Cel5a. As such, by increasing the thermostability of the Cel5a enzyme, the reaction temperature for cellulose digestion may be increased. Increasing the reaction temperature for cellulose digestion is useful because isolating cellulose from the lignocellulosic material requires heating at about 200° C. As such, being able to perform the subsequent cellulose hydrolysis step at temperatures higher than the current industry standard of about 50° C. may reduce the amount of energy necessary to cool the pretreated cellulose substrate. (Sun et al. 2002, *Bioresource Technol.*, 83:1-11, the entire contents of which are herein incorporated by reference.) This reduction in energy may result in a reduction in the overall production cost.

Some embodiments of the present invention include Cel5a variants having both increased thermostability and increased enzymatic activity compared to wild type (WT) Cel5a. As such, these Cel5a variants are able to digest cellulose at a higher temperature than wild type and digest more cellulose than wild type. For example, as disclosed herein, increased amounts of cellulase activity compared to WT Cel5a range from about 0.1 to about 100  $\mu$ M more cellobiose equivalents at 60° C., and from about 0.1 to about 1 mM (1,000  $\mu$ M) more cellobiose equivalents at 70° C.

Some embodiments of the present invention include Cel5a variants having increased expression compared to wild type. As such, more Cel5a protein may be isolated without an increase in the culture processing of the expression host organism. Some embodiments of the present invention include Cel5a variants having increased thermostability, increased activity, and increased expression compared to wild type Cel5a.

Abbreviations for amino acids are used throughout this disclosure and follow the standard nomenclature known in the art. Alanine is Ala or A; Arginine is Arg or R; Asparagine is Asn or N; Aspartic Acid is Asp or D; Cysteine is Cys or C; Glutamic acid is Glu or E; Glutamine is Gln or Q; Glycine is Gly or G; Histidine is His or H; Isoleucine is Ile or I; Leucine is Leu or L; Lysine is Lys or K; Methionine is Met or M; Phenylalanine is Phe or F; Proline is Pro or P; Serine is Ser or S; Theonine is Thr or T; Tryptophan is Trp or W; Tyrosine is Tyr or Y; and Valine is Val or V.

For point mutations, amino acid substitutions are shown, for example, as E53D, where the wild type glutamic acid (E) at position 53 is changed to aspartic acid (D). Also, S318P/Q/E denotes that the wild type serine (S) at position 318 may be changed to proline (P), glutamine (Q), or glutamic acid (E).

As used herein the terms “variant” and “mutant” are used interchangeably to refer to a protein having a change in the wild type protein sequence.

#### Expression of HjCel5a

The Cel5a variants of the present invention have at least one point mutation in the amino acid sequence of wild type *Hypocrea jecorina* (*Trichoderma reesei*) Cel5a. Wild type HjCel5a is encoded by the DNA sequence of SEQ ID NO: 1. As used herein, Cel5a and HjCel5a are used interchangeably to refer to the endoglucanase (SEQ ID NO: 3) encoded by SEQ ID NO: 1 or an indicated variant thereof.

The HjCel5a variant and wild type proteins as disclosed herein are transformed and expressed in the yeast host *Saccharomyces cerevisiae*. However, the Cel5a variants of the present invention may be expressed in any suitable host organism. Non-limiting examples of host organisms include *Trichoderma*, *Saccharomyces*, *Pichia* and *Aspergillus*. Any industrial expression system with known culturing methodologies may be used. Examples of known host organisms are described in Murray P, et al. 2004, *Protein Expr Purif*, 38:248-257; Qin et al. 2008, *Protein Expr Purif* 58:162-167; Celik et al., 2012, *Biotechnol. Adv.*, 30:1108-1118; and Su et al., 2012, *Adv. Appl. Microbiol.*, 81:1-61, the entire contents of all of which are herein incorporated by reference.

All expression plasmids encoding enzyme proteins disclosed herein, including Cel5a, may be constructed by any appropriate methods known in the art. Methods for constructing the various Cel5a plasmids are disclosed in the Examples, but the present invention is not limited by the particular plasmid sequences disclosed herein, excluding the indicated Cel5a variant sequence.

#### Cel5a Thermostability Mutants

Thermostability of Cel5a mutants was assayed using a T<sub>50</sub> or T<sub>450</sub> assay as described in Examples 1 and 4, respectively. The T<sub>50</sub> and T<sub>450</sub> assays determine the temperature at which half of the maximal protein activity remains, and is used to assay thermostability of a protein. As used herein, the temperatures determined by a T<sub>50</sub> or T<sub>450</sub> assay approximate each other, and are therefore used interchangeably. In some embodiments of the present invention, a Cel5a mutant endoglucanase having a T<sub>50</sub> above the T<sub>50</sub> of wild type Cel5a endoglucanase includes a point mutation in the wild type amino acid sequence of Cel5a. The Cel5a amino acid point mutations conferring an increased T<sub>50</sub> over wild type Cel5a include: D13E, T18P, E53D, T57N, N76P, S79P/Q, T80E, I82M, V101I, S133R, S139P, N153D, N155E/Q, T156E, G189S/A/E, F191V, K219Q/A, T233V, G239E/N/D, V265T, D271Y/F, Y278F, G293A, S309W/F/L, S318P/Q/E, or any combination thereof, with mutant residue numbering based on SEQ ID NO: 5.

Cel5a combination mutants of the listed point mutations above also confer increased thermostability compared to wild type Cel5a. In some embodiments a set of point mutations in Cel5a, with mutant residue numbering based on SEQ ID NO: 5, having increased thermostability over wild type Cel5a include one of the following sets:

N76P and S139P;  
T57N, N76P, S139P, and N155E;  
T57N, N76P, S139P, N155E, and G189S;  
T57N, N76P, S139P, N155E, and K219Q;  
T57N, N76P, S139P, N155E, and G239N;

## 11

T57N, N76P, S139P, N155E, and D271F;  
 T57N, N76P, S139P, N155E, D271F, and Y278F;  
 T57N, N76P, S139P, N155E, G189S, and K219Q;  
 T57N, N76P, S139P, N155E, G189S, and G239N;  
 T57N, N76P, S139P, N155E, G189S, K219Q, and G239N;  
 T57N, N76P, S139P, N155E, G189S, D271F, and Y278F;  
 T57N, N76P, S139P, N155E, G189S, D271F, Y278F, and  
 G293A;

T57N, N76P, T80E, S139P, N155E, G189S, D271F,  
 Y278F, and G293A;

T57N, N76P, S139P, N155E, G189S, K219Q, G239N,  
 D271F, Y278F, and G293A;

T57N, N76P, T80E, S139P, N155E, G189S, K219Q,  
 G239N, D271F, Y278F, and G293A;

T57N, N76P, T80E, S139P, N155E, G189S, K219Q,  
 G239N, D271F, Y278F, G293A, and S318P;

T57N, N76P, T80E, S139P, N155E, G189S, K219Q,  
 G239N, D271F, Y278F, G293A, S309F, and S318P;

T80E, S133R, N155E, G239E, Y278F, and S318Q;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309L, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309W, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309L, and S318E;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309W, and S318E;

G189A, F191V, G239D, S242A, V265T, D271Y, S318P/E,  
 and S322A; or

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 F191V, T233V, G239E, V265T, D271Y, G293A, S309W, and  
 S318P.

These point mutations and sets of point mutations are char-  
 acterized for thermostability as disclosed, for example, in  
 Examples 1-4 and FIGS. 1, 5, 7, 9B, 11A, 11B, 14A, 14B, 17,  
 and 19A.

Cel5A Mutants Having Increased Thermostability and Activ-  
 ity

The Cel5a mutants having increased thermostability were  
 assayed for enzymatic activity using, for example, a cellulase  
 activity assay measuring cellobiose digestion, as described in  
 Examples 1 and 6. The amino acid point mutations in Cel5a  
 that confer both an increased  $T_{50}$  and increased enzymatic  
 activity compared to wild type Cel5a include: E53D, T57N,  
 S79P, T80E, V101I, S133R, N153D, N155E, T156E, G189S,  
 F191V, T233V, G239E, V265T, D271Y, G293A, S309W/L,  
 or S318P/Q/E, with mutant residue numbering based on SEQ  
 ID NO: 5. These point mutations are characterized for ther-  
 mostability and activity, for example, in Examples 1-4, Tables  
 I-IV, and FIGS. 4, 10B, 10C, 16A, 16B, and 17.

In some embodiments of the present invention, combina-  
 tions of point mutations confer increased thermostability and  
 activity. In some embodiments, a Cel5a mutant endogluca-  
 nase having both increased thermostability and activity has a  
 combination of point mutations in the Cel5a amino acid  
 sequence including one of the following sets of mutations,  
 with mutant residue numbering based on SEQ ID NO: 5:

G189A, F191V, G239D, S242A, V265T, D271Y, S318P/E,  
 and S322A;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309L, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309W, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309L, and S318E;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309W, and S318E; or

## 12

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 F191V, T233V, G239E, V265T, D271Y, G293A, S309W, and  
 S318P.

The above combination mutants are characterized for ther-  
 mostability and activity, for example, in Examples 4-6, Tables  
 V and VI, and FIGS. 18A, 18B, 20C, 20D, 21C, 21D, 22A,  
 and 22B.

In some embodiments of the present invention, a Cel5a  
 mutant endoglucanase has increased thermostability (as mea-  
 sured by a  $T_{50}$  assay) up to 17° C. above the  $T_{50}$  of wild type  
 Cel5a and retains the same activity at least at 60° C.

Cel5A Mutants Having Increased Expression

Some Cel5a mutants confer an increase in expression as  
 compared to wild type Cel5a. As used herein, "protein expres-  
 sion" refers to the amount of protein produced in a host  
 organism. Cel5a is expressed into the medium of the host  
 cell—for example, *S. cerevisiae*. The single amino acid point  
 mutations in Cel5a that confer increased protein expression  
 of the Cel5a mutant protein in *S. cerevisiae* compared to the  
 level of wild type Cel5a protein expression in *S. cerevisiae*,  
 include: D13E, T18P, G64A, N76P, S79P/Q/E, T80E/Q,  
 I82L/M, V101I/L, A122E, S133R, S139P, N155E/Q, T156E,  
 G189S/A/E/K, K219Q/A, G239E/N/D, V265T, D271Y/F,  
 Y278F, G293A, S309W/F/L, or S318P/Q/E, with mutant  
 residue numbering based on SEQ ID NO: 5. These point  
 mutations are characterized for protein expression in, for  
 example, Examples 1-3 and Tables I-IV and VI.

In some embodiments of the present invention, combina-  
 tions of point mutations confer increased Cel5a protein  
 expression. In some embodiments a Cel5a mutant endogluca-  
 nase has a combination of point mutations in the Cel5a  
 amino acid sequence, with mutant residue numbering based  
 on SEQ ID NO: 5, including one of the following sets of  
 mutations:

T80E, S133R, N155E, G239E, Y278F, and S318Q;  
 G189A, F191V, G239D, S242A, V265T, D271Y, S318P/E,  
 and S322A;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309L, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309W, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309L, and S318E;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309W, and S318E; or

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 F191V, T233V, G239E, V265T, D271Y, G293A, S309W, and  
 S318P.

Excluding the combination set of (T80E, S133R, N155E,  
 G239E, Y278F, and S318Q), the Cel5a combination mutants  
 having increased Cel5a expression as listed above, also have  
 increased thermostability and enzymatic activity as disclosed  
 herein. As such, each of the combinations of point mutations  
 (excluding the indicated combination) produce a Cel5a  
 mutant that is stable at a higher temperature than wild type  
 Cel5a, is more enzymatically active than wild type Cel5a, and  
 is expressed at higher levels.

The following Examples are presented for illustrative pur-  
 poses only, and do not limit the scope or content of the present  
 application.

## EXAMPLE 1

Identification of D13E, E53D, T57N, G189A,  
 G293A Cel5a Mutants

Experiments were pursued to uncover stabilizing consen-  
 sus mutations in the primary endoglucanase Cel5A from

*Hypocrea jecorina* (HjCel5A), a molecule with ~400 homologous sequences in the NCBI non-redundant protein database. Using this data, six multiple sequence alignments (MSAs) were constructed varying in the number, level of characterization, and percent identity to the query of the aligned sequences. The alignments were filtered with numerical thresholds to reveal highly conserved residues (high relative entropy) at positions able to mutate independently of other protein sites (low mutual information). Using this method, five thermostabilizing point mutations were identified: D13E (+3.0° C.), E53D (+2.7° C.), T57N (+1.1° C.), G189A (+0.4° C.), and G293A (+3.6° C.), with mutant residue numbering based on SEQ ID NO: 5.

Experimental Screening and Validation of Thermostability Enhancing Mutations. After applying the relative entropy (RE) and mutual information (MI) constraints, a total of 21 unique mutations were predicted as stabilizing from six consensus sequence alignments. Point mutants were constructed, proteins were secreted from *Saccharomyces cerevisiae*, and supernatants were screened for activity on Avicel, a microcrystalline cellulose powder, at a temperature two degrees higher (73° C.) than the wild type (WT)  $T_m$  (FIG. 1). As a result of screening supernatant, high signal can indicate greater thermostability, activity, and/or expression, all desirable traits. Mutations D13E, E53D, T57N, I82L, V101L, G189A, and G293A demonstrated greater activity than WT and were selected for further characterization, with mutant residue numbering based on SEQ ID NO: 5.

The seven candidate HjCel5A point mutants were purified and pre-incubated at a gradient of temperatures for 10 minutes (min) before adding Avicel to assess activity over 2 hours at 60° C. Five mutations exhibited a  $T_{50}$ , the temperature at which half of the total enzyme remains active, greater than WT ( $\Delta T_{50, D13E}=3.0$ ,  $\Delta T_{50, E53D}=2.7$ ,  $\Delta T_{50, T57N}=1.1$ ,  $\Delta T_{50, G189A}=0.4$ , and  $\Delta T_{50, G293A}=3.6$ ° C.) (Table I, FIG. 2, FIGS. 3A-3G). The two remaining mutations exhibited slightly lower stabilities than WT ( $\Delta T_{50, I82L}=-0.3$ ,  $\Delta T_{50, V101L}=-0.4$ ) and likely exhibit high activity on the screen due to increases in expression level (Table I).

TABLE 1

Structural Analysis of Stabilizing Mutations, with mutant residue numbering based on SEQ ID NO: 5

Mutations	$T_{50, WT}^a$ (° C.)	$T_{50, mut}$ (° C.)	$\Delta T_{50}$ (° C.)	Activity ( $\mu$ M Cellobiose Equivalents)	$\Delta$ Activity ( $\mu$ M Cellobiose Equivalents)	Expression Level Relative to WT	Location
WT	—	—	—	193.7 $\pm$ 12.23	0.0	—	—
D13E	68.6 $\pm$ 0.3	71.5 $\pm$ 0.4	3.0 $\pm$ 0.5	184.4 $\pm$ 1.15	-9.3	1.6	Core
E53D	68.7 $\pm$ 0.3	71.4 $\pm$ 0.6	2.7 $\pm$ 0.7	201.9 $\pm$ 4.91	8.2	0.9	Boundary
T57N	71.0 $\pm$ 0.0	72.1 $\pm$ 0.0	1.1 $\pm$ 0.0	240.7 $\pm$ 4.08	47.0	0.3	Surface
G189A	70.8 $\pm$ 0.3	71.2 $\pm$ 0.3	0.4 $\pm$ 0.4	190.3 $\pm$ 3.37	-3.4	1.2	Boundary
G293A	70.7 $\pm$ 0.1	74.3 $\pm$ 0.1	3.9 $\pm$ 0.2	221.0 $\pm$ 1.75	27.3	0.8	Core
I82L	69.1 $\pm$ 0.4	68.9 $\pm$ 0.3	-0.2 $\pm$ 0.5	N/A	N/A	1.6	Core
V101L	68.8 $\pm$ 0.1	68.3 $\pm$ 0.3	-0.5 $\pm$ 0.3	N/A	N/A	2.3	Core

<sup>a</sup>The  $T_{50}$  of WT HjCel5A fluctuates by 1° C. due to variables described in the materials and methods section. All mutants were assayed simultaneously with a WT standard.

Activity of Stabilizing Mutations. To ensure that the stabilizing mutations do not adversely affect enzymatic activity, the five point mutants were tested for hydrolysis on Avicel after 2 hours at 60° C. (FIG. 4). The T57N and G293A mutants show significantly elevated activity, while the activities of the remaining mutants are comparable to WT (Table I).

MSA Construction and Analysis. Sequences homologous to the catalytic domain of HjCel5A (from GVR to CLARKG) were retrieved using the Position-Specific Iterated BLAST (PSI-BLAST) database search applied to the non-redundant protein sequences National Center for Biotechnology Infor-

mation (NCBI) database. (Altschul et al. 1997, *Nucleic Acids Research*, 25:3389-3402, the entire contents of which are herein incorporated by reference.) Constraints on the percent identity of the sequences to the query were introduced using the formatting options feature within the BLAST tool. Relative entropy was calculated using the yeast codon probabilities from Sullivan et al. 2012, *J. Molec. Biol.*, 420:384-399, the entire contents of which are herein incorporated by reference.

Cel5A Plasmid Construction. The Cel5A gene was synthesized by DNA 2.0 (Menlo Park, Calif., USA) with codon frequency optimized for *S. cerevisiae*. The construct consists of an  $\alpha$ MFpp8 secretory leader sequence (GenBank BK006949 193648-194145) followed by a region coding for the CBM from the *H. jecorina* CBM (GenBank ABA64553.1) preceded by an extra 'AR' introduced during cloning. This DNA sequence (SEQ ID NO: 2) is:

```
5' -GGCTAGACAACAAACAGTATGGGGTCAATGTGGTGGTATTGGAT
GGTCTGGTCCGA -CAAACGTGTCTCCAGGCTCGCATGTTTCGACACT
AAATCCATATTACGCTCAATGTATCCCTGGCGCTACCACTATAACAA
CTTCTACTAGACCACCTTCTGGTCCGACGACAACTACAAGGGCTACC
TCAACCTCTTCTCTACACCCCTACTTCCAGC-3'
```

The additional 'AR' sequence does not significantly affect any protein properties. The CBM region is then followed by an HjCel5A catalytic domain sequence identical to GenBank entry JN172972.1. This construct contains a short linker and an N-terminal His-tag. The assembled gene was cloned into the yeast expression vector YEp352/PGK91-1- $\alpha$ ss between the BglIII and MboI restriction sites using the Gibson assembly method, as described in Gibson et al. 2009, *Nature Methods*, 6:343-345, the entire contents of which are herein incorporated by reference. Point mutations were introduced using

the QuikChange Lightning Site-Directed Mutagenesis Kit from Agilent Technologies (Santa Clara, Calif., USA) using primers designed with the online tool provided by Agilent:

Following sequence verification, clones were transformed into YDR483W BY4742  $\Delta$ Kre2 *S. cerevisiae* cells using the method outlined in Schiestl et al. "Introducing DNA into yeast by transformation," In: Johnston, M, Fields S., Eds., 1993, *Methods: A Companion to Methods in Enzymology*, Acad. Press, Inc. pp. 79-85, the entire contents of which are herein incorporated by reference.

## 15

Thermostability/Activity Screen. *S. cerevisiae* carrying the Hjc5A plasmid were inoculated into 1 mL SD-Ura media in 24-well plates and allowed to grow overnight at 30° C. with shaking at 200 rpm. 4 mL of YPD (yeast extract peptone-dextrose) were added and the cells were allowed to shake at 30° C. for an additional 48 hours before harvesting the supernatant through centrifugation. 5 µL of supernatant, 45 µL of YPD, and 60 µL of a 1.5% Avicel PH-101 (Sigma-Aldrich) slurry in 50 mL sodium acetate, pH 5.0 (cellulase buffer) were combined in a 96-well PCR plate and incubated for 1.5 hours to allow the CBM to bind to the substrate. The bound enzymes were washed three times with cellulase buffer and incubated at 73° C. for 2 hours. Following hydrolysis, 50 µL of the reaction supernatants were tested for reducing sugar concentrations via a modified Park-Johnson assay. All screen samples were run in duplicate.

Park-Johnson Assay. To detect reducing end release, 50 µL of sample were combined with 100 µL of reagent A (0.5 g L<sup>-1</sup> K<sub>3</sub>Fe<sub>3</sub>(CN)<sub>6</sub>, 34.84 g L<sup>-1</sup> PO<sub>4</sub>K<sub>2</sub>H, pH 6.0) and 50 µL of reagent B (5.3 g L<sup>-1</sup> Na<sub>2</sub>CO<sub>3</sub>, 0.65 g L<sup>-1</sup> KCN). In experiments resulting in high amounts of reducing ends, 25 µL of sample is combined with 175 µL of the 2A:1B mixture. After incubating the mixture at 95° C. for 15 minutes in a PCR block, the plate is cooled on ice for five minutes. In a flat well plate, 90 µL of reagent C (2.5 g L<sup>-1</sup> FeCl<sub>3</sub>, 10 g L<sup>-1</sup> polyvinyl pyrrolidone, 2 N H<sub>2</sub>SO<sub>4</sub>) is combined with 180 µL of the heat treated sample. The sample is then allowed to incubate for five minutes before measuring absorbance at 595 nm.

Enzyme Purification. Yeast colonies carrying the Hjc5A plasmid were inoculated into 6 mL of SD-Ura media and grown at 30° C. with shaking at 200 rpm. The preculture was then added to YPD and incubated for 48 hours. Following centrifugation, the supernatant was subjected to an 80% ammonium sulfate precipitation. The mixture was spun for 20 minutes at 8 kg and the pellet resuspended in 20 mL of lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.4, 300 mM NaCl, 10 mM imidazole). Following a pH adjustment to 7.4, the protein was rotated at 4° C. with 1 mL of Ni-NTA resin (Qiagen) conditioned with lysis buffer for 1 hour. The mixture was loaded into a gravity column, washed with 20 mL of lysis buffer, 20 mL of wash buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.4, 300 mM NaCl, 20 mM imidazole), and eluted with 6 mL of elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.4, 300 mM NaCl, 250 mM imidazole). After concentrating the elution to 0.5 mL, the protein was further purified and buffer exchanged into cellulase buffer through size exclusion chromatography. Protein concentrations were determined through measuring absorbance at 280 nm ( $\epsilon_{280}=81950 \text{ cm}^{-1} \text{ M}^{-1}$ ).

T<sub>50</sub> Assay. To assess thermostability via enzymatic activity, 40 µL of protein at a concentration of 0.25 µM was added to a PCR plate in triplicate for each of 12 temperatures. Enzyme was pretreated from 60-80° C. for ten minutes, then allowed to cool for an additional five minutes. 60 µL of a 1.5% Avicel slurry in cellulase buffer was added to each well and the plates were incubated at 60° C. for an hour. The plates were promptly cooled for 5 minutes on ice then centrifuged for 5 minutes to pellet the Avicel. Activity assessment with the

## 16

Park-Johnson assay immediately followed using a 50 µL sample volume. To compare T<sub>50</sub> values, the data were scaled from 0 to 1 using the following equation:

$$\text{Fraction Active} = \frac{(A_T - A_{min})}{(A_{max} - A_{min})}$$

In the above equation, A<sub>T</sub> is the activity as measured by A<sub>595</sub> at a particular temperature, A<sub>min</sub> is the lowest observed activity, and A<sub>max</sub> is the highest observed activity for a particular protein. T<sub>50</sub> values were derived from generating curve fits using the Hill equation:

$$\text{Curve Fit} = \frac{T^n}{T^n + m^n}$$

Here n is the Hill coefficient, m is the T<sub>50</sub>, and T is the temperature. Values for n and m were solved using the curve fit tool in MATLAB. (MATLAB and Statistics Toolbox Release 2011, The MathWorks, Inc., Natick, Mass., United States, the entire contents of which are herein incorporated by reference.) Because the T<sub>50</sub> can fluctuate by approximately 1° C. depending on fluctuations in Avicel milling, subtle changes in cooling time, and PCR plate edge effects, all samples were run simultaneously with a WT standard. The ΔT<sub>50</sub> values are calculated as T<sub>50, mut</sub> - T<sub>50, WT</sub>.

Single-Point Activity Assay. To determine enzyme activity, 40 µL of enzyme at 0.5 µM was combined with 60 µL of 1.5% Avicel in a PCR plate. The mixture was incubated at 60° C. for two hours to allow hydrolysis to proceed. After cooling the plate on ice for 5 minutes, 100 µL of 0, 50, 100, 150, 200, 250, 300, and 350 µM cellobiose standards were added to the plate in triplicate. The plate was centrifuged to pellet the Avicel and 25 µL of the samples were extracted to perform a Park Johnson activity assay. All samples were tested in triplicate.

## EXAMPLE 2

Identification of S79Q, T80E, I82M, V101I, S133R, N155E, N155O, T156E, G239E, Y278F, S318E, and S318O Cel5a Mutants

Utilizing a core repacking algorithm, 32 Hjc5A mutations were identified, with mutant residue numbering based on SEQ ID NO: 5. These 32 mutations were individually cloned as point mutations and screened for stability and adequate activity (FIG. 5), following methods disclosed in Example 1. Supernatants harboring secreted protein were screened for activity after incubation for 1 hour at 73° C., 3.5 degrees higher than the WT T<sub>m</sub> (69.5° C.). Only two mutations with activities exceeding WT were detected from the screening step—I82M and V101I. Two variants, each carrying one of these stabilizing mutations were expressed and purified to assess thermostability more directly. Both mutants showed a slight increase in thermostability with I82M conferring a 0.3 and V101I a 0.5° C. increase in T<sub>50</sub>, the temperature at which half of the maximal protein activity persists (Table II, FIGS. 6A, 6B).

TABLE II

Characterization of stabilizing mutants, with mutant residue numbering based on SEQ ID NO: 5								
Construct	T <sub>50,WT</sub> (° C.)	T <sub>50,mnt</sub> (° C.)	ΔT <sub>50</sub> (° C.)	Activity (μM Cellobiose Equivalents)	ΔActivity (μM Cellobiose Equivalents)	ΔΔG <sup>b</sup> (kcal mol <sup>-1</sup> )	Site MT <sup>b</sup>	Expression Level Relative to WT
WT	—	—	—	193.7 ± 12.2 Core Mutations	0	—	—	—
I82M	69.5 ± 0.2	69.8 ± 0.5	0.3 ± 0.5	188.5 ± 2.4	-5.2	-0.9	0.48	1.3
V101I	69.6 ± 0.3	70.1 ± 0.2	0.5 ± 0.4	210.1 ± 1.8 Helix Mutations	16.4	-0.4	0.29	1.7
T80E	69.3 ± 0.2	69.8 ± 0.1	0.5 ± 0.2	203.6 ± 9.2	9.8	2.7	0.73	2.3
S133R	68.9 ± 0.1	69.4 ± 0.1	0.4 ± 0.2	197.1 ± 2.6	3.4	1.8	0.88	1.8
N155E	69.5 ± 0.3	70.0 ± 0.1	0.5 ± 0.3	199.4 ± 1.1	5.6	0.5	0.69	4.9
N155Q	68.4 ± 0.1	68.5 ± 0.1	0.1 ± 0.1	172.8 ± 4.0	-20.9	0.1	0.69	1.1
T156E	69.5 ± 0.2	69.7 ± 0.3	0.2 ± 0.3	217.6 ± 7.9	23.9	1.2	0.98	4.9
G239E	69.7 ± 0.1	70.0 ± 0.3	0.2 ± 0.3	216.9 ± 6.5	23.2	-0.2	0.96	1.0
Y278F	69.2 ± 0.2	70.2 ± 0.4	1.0 ± 0.5	174.7 ± 2.7	-19.1	1.3	0.64	0.4
S318E	69.7 ± 0.2	70.5 ± 0.1	0.9 ± 0.2	244.1 ± 4.3	50.4	-0.5	N/A <sup>a</sup>	0.6
S318Q	68.9 ± 0.1	69.4 ± 0.2	0.5 ± 0.2	196.0 ± 2.4	2.3	-0.3	N/A <sup>a</sup>	0.9
S79Q	69.4 ± 0.2	69.5 ± 0.2	0.0 ± 0.3	174.4 ± 5.1	-19.3	0.4	0.61	1.4
S79E	69.9 ± 0.2	69.7 ± 0.0	-0.1 ± 0.2	N/A	N/A	-0.3	0.61	5.5
T80Q	69.2 ± 0.2	69.1 ± 0.1	-0.1 ± 0.2	N/A	N/A	3.2	0.73	2.0
A122E	69.0 ± 0.5	68.8 ± 0.2	-0.2 ± 0.5	N/A	N/A	1.0	0.51	3.2
G239Q	69.1 ± 0.1	68.5 ± 0.1	-0.9 ± 0.2	N/A	N/A	-1.6	0.96	0.9
Helix Combo	69.5 ± 0.5	71.9 ± 0.3	2.4 ± 0.5	N/A	N/A	N/A	N/A	4.5

<sup>a</sup>Insufficient homologous sequence data. <sup>b</sup>Values calculated from the 444 sequence MSA of Example 1.)

A helix dipole stabilization calculation as developed by Marshall et al. was used to select 44 predicted helix mutations. (Marshall et al. 2002, *J. of Molec. Biol.*, 316:189-199, the entire contents of which are herein incorporated by reference.) These 44 predicted helix mutations were constructed as point mutants and screened for activity (FIG. 7) with the same procedure used to probe core mutations. In the activity screen, 14 constructs demonstrated greater activity than WT. Following purification and activity screening at a gradient of temperatures, nine constructs demonstrated a positive ΔT<sub>50</sub> (T80E, S133R, N155E, N155Q, T165E, G239E, Y278F, S318E, and S318Q), four showed a decrease in thermostability (S79E, T80Q, A122E, and G239Q), and one behaved similarly to WT (S79Q) (Table II, FIGS. 8A-8N), with mutant residue numbering based on SEQ ID NO: 5. As is the case with the core mutations, the helix dipole stabilizing mutations provide modest stability benefits (ΔT<sub>50</sub> ≤ 1° C.). Only five of the nine stabilizing mutations exhibit a ΔT<sub>50</sub> ≥ 0.5° C.

Given the relatively low enhancements in thermostability observed for the helix dipole stabilizing mutations, it remained unclear whether these mutations would provide any tangible benefit. A combination construct (Helix combo) containing the T80E, S133R, N155E, G239E, Y278F, and S318Q mutations, with mutant residue numbering based on SEQ ID NO: 5, was constructed and the ΔT<sub>50</sub> (2.4° C.) and optimal reaction temperature (T<sub>opt, helix combo</sub> = 66° C., T<sub>opt, WT</sub> = 63.5° C.) were determined (FIGS. 9A-9B). Both values show modest increases of ~2.5° C. While the six stabilizing mutations did not additively increase T<sub>50</sub> and T<sub>opt</sub>, the combination mutant still demonstrates improved thermostability compared to the most beneficial helix point mutant. In addition, the combination mutant shows a 4.5 fold improvement in expression over WT (Table II).

Activity of Stabilized Mutants. Useful enzyme mutations not only confer stability, but also elevate or preserve activity. The activity of each stabilizing point mutation was tested at 60° C. for two hours on Avicel, a crystalline cellulose powder (Table II). While the two stabilizing core mutants have activi-

ties comparable to WT (FIGS. 10A, 10B), the helix mutations show an even distribution between lower and higher activities (FIG. 10C).

Expression. In addition to preserving activity, desirable mutations will also maintain or enhance protein expression levels. Five of the helix dipole stabilizing hits on the activity screen were either neutral or destabilizing. Four of these mutations confer greater protein expression than WT in *S. cerevisiae*, which is the probable cause for their high activity in the screen. In general, the point mutants and helix combination mutant demonstrated large increases in expression level (Table II). The catalytic domain of H<sub>3</sub>Cel5A V101I was expressed in *Escherichia coli*, resulting in a three-fold increase in protein yield.

Structure preparation. Designs were performed on chain A of the H<sub>3</sub>Cel5A crystal structure (PDB ID 3QR3) (SEQ ID NO:5). (Lee et al. 2011, *Protein Science*, 20:1935-1940, the entire contents of which are herein incorporated by reference.) After removing water molecules and ions, hydrogens were added to the structure using the protein process application within the design software TRIAD (Triad, 2012, Protabit, LLC, Pasadena, Calif.). This application was additionally employed to optimize the structure through 50 steps of gradient-based energy minimization using the energy function described in the computational design section.

Computational Design. Computational design parameters outlined in this section were kept consistent between the core repacking and helix dipole stabilization calculations. All calculations were executed using an energy function based on the DREIDING forcefield that includes terms for van der Waals hydrogen bonding, electrostatics, implicit solvation, and phi-psi propensities. (Mayo et al. 1990, *J. of Phys. Chem.*, 94:8897-8909; Dahiyat et al. 1997, *PNAS*, 94:10172-10177; and Dahiyat et al. 1997, *Prot. Science*, 6:1333-1337, the entire contents of all of which are herein incorporated by reference.) In calculating the implicit solvation term, an occlusion-based solvation potential was applied with scale factors of 0.05 for nonpolar burial, 2.5 for nonpolar exposure, and 1.0 for polar burial, as described in Chica et al. 2010,



*PNAS*, 107:20257-20262, the entire contents of which are herein incorporated by reference. Sequence optimization was performed with FASTER and a Monte Carlo-based algorithm was used to sample sequences near the minimum energy sequence. (Allen et al. 2006, *J. Computational Chem.*, 27:1071-1075; Desmet et al. 2002, *Proteins: Structure, Function, and Bioinformatics*, 48:31-43; Metropolis et al. 1953, *J. of Chem Physics*, 21:1087-1092; and Voigt et al. 2000, *J. of Molec. Biol.*, 299:789-803, the entire contents of all of which are herein incorporated by reference.

**Cel5A Plasmid Construction.** Cel5a mutants were constructed as described in Example 1.

**Thermostability/Activity Screen.** Thermostability and activity screens were performed as described in Example 1.

**$T_{opt}$  Assay.** To assess the optimal operating temperature of HjCel5A constructs, 40  $\mu$ L of protein at a concentration of 0.25  $\mu$ M was combined with 60  $\mu$ L of a 1.5% Avicel slurry in cellulase buffer in a PCR plate in triplicate for each of 12 temperatures. The plates were incubated at 60° C. for two hours and promptly cooled for 5 minutes on ice. After centrifugation for 5 minutes to pellet the insoluble substrate, activity was assessed with the Park-Johnson assay as described in Example 1, using a 25  $\mu$ L sample volume. Bovine serum albumin (BSA) at a final concentration identical to the protein of interest served as a negative control.

**Enzyme Purification.** Cel5a proteins were cultured and purified as described in Example 1.

**$T_{50}$  Assay.** The T50 assay was performed as described in Example 1.

**Single-Point Activity Assay.** The single-point activity assay was performed as described in Example 1.

mutations, FoldX predicted 1008 as stabilizing ( $\Delta\Delta G < 0$  kcal mol<sup>-1</sup>). As this sizeable pool of candidate mutations was too large to screen using available resources, mutations with a  $\Delta\Delta G \leq -1.75$  kcal mol<sup>-1</sup> were examined. In addition, the energy cutoff was tightened from -0.75 to -1.75 kcal mol<sup>-1</sup> as described in Komor et al. 2012, *Prot. Eng. Des. Sel.*, 25:827-833, the entire contents of which is herein incorporated by reference. Applying this stringent criterion reduced the candidate pool to 43 mutations.

**Triad  $\Delta\Delta G$  Designs.** In parallel to the FoldX designs, site-saturated mutagenesis using Triad was used with the Rosetta forcefield and the HjCel5A crystal structure. Out of 6,232 possible point mutations, 789 were predicted as stabilizing ( $\Delta\Delta G < 0$ ). To reduce the number of candidate mutations to a manageable quantity, an arbitrary -1.75 kcal mol<sup>-1</sup> cutoff was applied leaving 47 mutations.

**Detecting Stabilizing Mutations.** Point mutants for each of the 43 FoldX and 44 Triad mutations were constructed for secretion in *Saccharomyces cerevisiae*. Proteins were expressed and the supernatant was screened for activity at 73° C., 3.5 degrees higher than the melting temperature ( $T_m$ ) of the native protein. Six FoldX and five Triad mutants demonstrated higher activity than wild type (WT) and were selected for more rigorous characterization.

As the screen was performed on unpurified protein in supernatant, hits may indicate a variant with improved thermostability, yield, activity, or a combination thereof. To determine the source of their improvement, the mutants and WT were simultaneously assayed for activity at 60° C. following a 10 minute incubation at a gradient of temperatures ranging from 60-80° C. The  $T_{50}$ , the temperature at which half of the maximum protein activity persists, of the mutation was computed and compared to the WT value. All of the mutations proved more thermostable than WT ( $\Delta T_{50} > 0^\circ$  C.), as summarized in Table III below, with mutant residue numbering based on SEQ ID NO: 5.

TABLE III

Characterization of FoldX and Triad $\Delta\Delta G$ mutations									
Construct	$T_{50,WT}$ (° C.)	$T_{50,mu}$ (° C.)	$\Delta T_{50}$ (° C.)	Activity ( $\mu$ M Cellobiose Equivalents)	$\Delta$ Activity ( $\mu$ M Cellobiose Equivalents)	FoldX $\Delta\Delta G$ (kcal mol <sup>-1</sup> )	Triad $\Delta\Delta G$ (kcal mol <sup>-1</sup> )	Expression Level (Mut/WT)	Location
WT	—	—	—	193.7 $\pm$ 12.2	0.0	—	—	—	
FoldX									
S79P	69.9 $\pm$ 0.3	70.2 $\pm$ 0.5	0.3 $\pm$ 0.5	205.1 $\pm$ 6.7	11.4	-1.84	0.97	1.0	Surface
N153D	70.3 $\pm$ 0.7	70.7 $\pm$ 0.6	0.5 $\pm$ 0.9	196.2 $\pm$ 1.5	2.5	-2.91	-0.25	0.2	Surface
D271F	70.5 $\pm$ 0.9	73.6 $\pm$ 0.6	3.1 $\pm$ 1.1	167.2 $\pm$ 6.3	-26.5	-2.23	-1.08	0.1	Boundary
D271Y	71.3 $\pm$ 0.3	73.9 $\pm$ 0.1	2.7 $\pm$ 0.4	209.5 $\pm$ 3.9	32.5	-1.83	-1.07	0.4	Boundary
S309L	71.3 $\pm$ 0.3	72.7 $\pm$ 0.3	1.5 $\pm$ 0.3	196.1 $\pm$ 1.9	2.4	-1.78	-2.26	0.8	Boundary
S318P	69.9 $\pm$ 0.6	73.1 $\pm$ 0.6	3.2 $\pm$ 0.9	218.5 $\pm$ 3.4	24.8	-2.32	3.27	0.3	Surface
Triad									
K219A	68.1 $\pm$ 0.7	70.1 $\pm$ 0.1	2.0 $\pm$ 0.7	167.0 $\pm$ 6.2	-26.7	2.07	-3.20	1.4	Core
K219Q	68.5 $\pm$ 0.0	71.3 $\pm$ 0.1	2.8 $\pm$ 0.1	162.0 $\pm$ 1.4	-31.7	1.28	-1.92	1.2	Core
S309F	68.3 $\pm$ 0.1	71.0 $\pm$ 0.1	2.7 $\pm$ 0.1	164.1 $\pm$ 8.6	-29.7	-0.57	-1.94	0.8	Boundary
S309L	71.3 $\pm$ 0.3	72.7 $\pm$ 0.3	1.5 $\pm$ 0.3	196.1 $\pm$ 1.9	2.4	-1.78	-2.26	0.8	Boundary
S309W	68.2 $\pm$ 0.0	68.6 $\pm$ 0.1	0.4 $\pm$ 0.1	202.8 $\pm$ 6.8	9.1	0.19	-2.24	0.5	Boundary

## EXAMPLE 3

Identification of T18P, N76P, S79P, S139P, N153D, G189A, G189S, G189E, G189Q, K219A, K219Q, G239S, G239D, G239N, D271Y, D271F, G293A, S309L, S309F, S309W, S318P, Cel5a Mutants

**FoldX Designs.** In silico site-saturated mutagenesis was performed on chain A from the HjCel5A crystal structure (PDB ID 3QR3) (SEQ ID NO:5). Out of 6,232 possible point

FoldX predicted three highly stabilizing mutations, D271F ( $\Delta T_{50} = 3.1^\circ$  C.), D271Y ( $\Delta T_{50} = 2.7^\circ$  C.), and S318P ( $\Delta T_{50} = 3.2^\circ$  C.), with mutant residue numbering based on SEQ ID NO: 5. Similarly, three mutations among the Triad predictions, K219A ( $\Delta T_{50} = 2.0^\circ$  C.), K219Q ( $\Delta T_{50} = 2.8^\circ$  C.), and S309F ( $\Delta T_{50} = 2.7^\circ$  C.), conferred significant increases in stability, with mutant residue numbering based on SEQ ID NO: 5. The mutant expression levels were similar or lower than WT.

Activity of Stabilizing FoldX and Triad Mutants. To determine whether the stabilizing FoldX and Triad mutations impact activity, purified protein was assayed for activity after 2 hours at 60° C. (Table III, FIGS. 11A, 11B and 13A, 13B). While the Rosetta mutations show decreased activity compared to WT, the FoldX mutations exhibit more diversity in activity. Notably, the FoldX mutations S318P and D271Y elevate the WT activity by 12.8 and 16.8%.

Backbone Stabilization: Removing Glycines and Adding Prolines. To test the hypothesis that adapting empirical methods to improve backbone stability may provide additional stabilizing mutations, all Gly  $\rightarrow$  X<sub>AA</sub> and X<sub>AA</sub>  $\rightarrow$  Pro mutations were fetched and ranked by  $\Delta\Delta G$  value. All of the mutations with  $\Delta\Delta G$  values  $\leq 0$  kcal mol<sup>-1</sup> were designated as potentially stabilizing. This relaxed cutoff allowed for the prediction of mutations that did not pass the -1.75 kcal mol<sup>-1</sup> threshold enforced in the general Triad  $\Delta\Delta G$  calculation. Only five mutual members appear in both the general Triad and glycine mutation lists. In addition, all predicted proline mutations were previously uncharacterized.

In total, 51 glycine and 46 proline mutations were predicted as stabilizing. Due to screening constraints, only the top 44 glycine (FIG. 14A) and 46 proline (FIG. 14B) mutants were constructed and screened using the setup described for the FoldX and Rosetta constructs. Nine glycine and three proline mutants demonstrated higher activity than WT on the screen. After purifying these enzymes and determining their  $\Delta T_{50}$ s, five glycine (G189A  $\Delta T_{50}$ =0.4° C., G189S  $\Delta T_{50}$ =1.2° C., G239D  $\Delta T_{50}$ =0.4° C., G239N  $\Delta T_{50}$ =0.7° C., and G293A  $\Delta T_{50}$ =3.5° C.) and three proline mutations (T18P  $\Delta T_{50}$ =2.0° C., N76P  $\Delta T_{50}$ =2.0° C., and S139P  $\Delta T_{50}$ =2.0° C.) demonstrated thermostability benefits (Table IV, FIGS. 15A-15B, 16A-16B), with mutant residue numbering based on SEQ ID NO: 5.

optimize and repair PDB functions within the software. A position scan was performed to compute energy values for WT and mutations to all other 19 amino acids. To compute  $\Delta\Delta G$  values, each mutation was compared to WT using the following equation:

$$\Delta\Delta G_{Mut} = \Delta G_{Mut} - \Delta G_{WT}$$

where  $\Delta G_{mut}$  is the energy computed for the mutation and  $\Delta G_{WT}$  is the energy computed for the WT residue at the same position. All calculations were performed using default parameters unless otherwise specified.

Triad  $\Delta\Delta G$  Calculation. All Rosetta calculations were performed using a modified version of the rosetta energy function described by Rohl et al. and implemented within the protein design software Triad (2012, Protabit, LLC, Pasadena, Calif.). (Rohl et al. 2004, *Methods in Enzymol.*, Academic press, pp. 66-93, the entire contents of which are herein incorporated by reference. The version of rosetta implemented in Triad employs a softer Lennard-Jones potential, a different set of amino-acid reference energies, and modified hydrogen bond and amino acid propensity weights. The energy function also lacks terms unnecessary for point mutation calculations including those for disulfide bonding, Ramachandran, proline closure, and omega tethering. Designs were performed on chain A of the HJcel5A crystal structure (PDB ID 3QR3)(SEQ ID NO: 5) (Lee et al. 2011, supra). After removing water molecules and ions, hydrogens were added to the structure using the protein process application within the design software Triad. Triad was additionally employed to optimize the structure through 50 steps of gradient-based energy minimization using the rosetta forcefield.

Glycine Scan. The glycine scan was performed in Triad using the modified version of the rosetta forcefield described in the Triad  $\Delta\Delta G$  Calculation section (above). The scan

TABLE IV

Backbone stabilizing mutations									
Construct	$T_{50,WT}$ (° C.)	$T_{50,Mut}$ (° C.)	$\Delta T_{50}$ (° C.)	Activity ( $\mu$ M Cellobiose Equivalents)	$\Delta$ Activity ( $\mu$ M Cellobiose Equivalents)	FoldX $\Delta\Delta G$ (kcal mol <sup>-1</sup> )	Triad $\Delta\Delta G$ (kcal mol <sup>-1</sup> )	Expression Level (Mut/WT)	Location
WT	—	—	—	193.7 $\pm$ 12.2	0	—	—	—	
GLY $\rightarrow$ X <sub>AA</sub>									
G189A	70.8 $\pm$ 0.3	71.2 $\pm$ 0.3	0.4 $\pm$ 0.4	170.1 $\pm$ 3.4	-23.7	-0.93	-0.76	1.2	Boundary
G189S	70.1 $\pm$ 0.3	71.3 $\pm$ 0.2	1.2 $\pm$ 0.4	208.2 $\pm$ 6.0	14.5	-0.45	-0.97	1.2	Boundary
G239D	70.8 $\pm$ 0.1	71.2 $\pm$ 0.1	0.4 $\pm$ 0.2	185.6 $\pm$ 3.1	-8.1	-0.80	-1.46	1.3	Boundary
G239N	70.4 $\pm$ 0.1	71.1 $\pm$ 0.0	0.7 $\pm$ 0.1	174.9 $\pm$ 4.3	-18.8	-0.13	-1.43	1.1	Boundary
G293A	70.3 $\pm$ 0.1	73.7 $\pm$ 0.1	3.5 $\pm$ 0.2	221.0 $\pm$ 1.2	27.3	6.66	-0.08	0.8	Core
G189E	70.6 $\pm$ 0.1	70.6 $\pm$ 0.2	0.0 $\pm$ 0.2	N/A	N/A	-0.58	-1.01	2.0	Boundary
G64A	70.3 $\pm$ 0.1	69.6 $\pm$ 0.1	-0.7 $\pm$ 0.1	190.5 $\pm$ 5.0	-3.2	2.97	-0.30	2.3	Core
G189K	70.6 $\pm$ 0.1	70.5 $\pm$ 0.1	-0.1 $\pm$ 0.2	N/A	N/A	-0.89	-0.55	1.2	Boundary
G239S	70.8 $\pm$ 0.2	69.7 $\pm$ 0.3	-1.0 $\pm$ 0.3	N/A	N/A	-0.20	-1.04	0.9	Boundary
X <sub>AA</sub> $\rightarrow$ PRO									
T18P	70.2 $\pm$ 0.1	70.4 $\pm$ 0.0	0.2 $\pm$ 0.1	187.5 $\pm$ 3.8	-6.2	-1.67	0.46	1.1	Surface
N76P	69.7 $\pm$ 0.2	70.5 $\pm$ 0.4	0.8 $\pm$ 0.5	177.3 $\pm$ 6.8	-16.4	0.00	0.95	1.7	Surface
S139P	69.6 $\pm$ 0.2	71.5 $\pm$ 0.6	1.8 $\pm$ 0.6	152.1 $\pm$ 8.0	-41.6	-1.33	2.14	1.8	Surface
Disulfide Mutations									
I44C/G91C	69.8 $\pm$ 0.4	69.3 $\pm$ 0.5	-0.5 $\pm$ 0.7	N/A	N/A	N/A	N/A	0.8	Surface/ Boundary

FoldX Calculation. All FoldX calculations were performed with version 3.0, as described in Schymkowitz et al. 2005, *Nucl. Acids Research*, 33:W382-W388, the entire contents of which are herein incorporated by reference. After removing waters and ligands, Chain A from the HJcel5A crystal structure (PDB ID: 3QR3) (SEQ ID NO:5) was prepared using the

mutates glycine in the native structure to each of the other 19 amino acids. All other conditions addressed in the rosetta calculation section apply to the glycine scan. The information calculated here is simply a subset of the data retrieved from the more comprehensive Triad-rosetta  $\Delta\Delta G$  scan and is reformatted to facilitate data analysis.

Proline Scan. The proline scans were performed with a restricted version of the algorithm used for the original Triad  $\Delta\Delta G$  scan calculation. This scheme calculates  $\Delta\Delta G$  values for mutating every position in the protein to proline and provides a ranked list of mutations. As in the glycine calculation, the generated information is a reformatted subset of the data retrieved from the Triad  $\Delta\Delta G$  scan.

Disulfide Bond Engineering Calculations. Disulfide bond engineering calculations were performed using the ssdesign application in the protein design software Triad. If two cysteine (CSS) rotamers come in close contact, the program adopts smaller values for force constants and barriers for DREIDING bonds, angles, and torsions. This leniency has been optimized to detect native disulfides as many disulfide geometries show slight deviations from canonical values. To design disulfides, the rotamer optimization algorithm simultaneously switches a pair of residues to CSS rotamers. Pair moves are biased towards those with good pairwise energies, i.e., those likely to form disulfides. Calculations were performed with 7 trajectories, a rotamer pair factor of 10, an iterations multiplier of 5, a disulfide force constant of 15, a disulfide max benefit of 35, and a CSS penalty of 15. All calculations employed a version of the rosetta forcefield, as described in the Triad  $\Delta\Delta G$  Calculation section (above), modified to employ the DREIDING disulfide bonding energy terms. The term used for bonds within disulfides is:

$$E_{bond}=12700.0B_{1,2}(r_{1,2}-R_{1,2})^2$$

where B is the bond order (1, 1.5, 2, or 3), r is the Cartesian distance between atoms 1 and 2, and  $R_{1,2}$  is the equilibrium bond distance between atoms 1 and 2. The term for disulfide angles is as follows:

$$E_{angle}=12100.0(\angle_{1,2,3}-\theta_{1,2,3})^2$$

where  $\angle_{1,2,3}$  is the observed angle between atoms 1, 2, and 3 and  $\theta_{1,2,3}$  is the equilibrium angle. Disulfide torsion angles are defined as:

$$E_{torsion}=12K_{1,2,3,4}N(1-d_{1,2,3,4}\cos(n_{1,2,3,4}\chi_{1,2,3,4}))$$

where  $K_{1,2,3,4}$  is the energy barrier, N is the number of torsion terms where atoms 2 and 3 are placed in the center,  $d_{1,2,3,4}$  is the phase factor (1 for cis, -1 for trans),  $n_{1,2,3,4}$  is the periodicity, and  $\chi_{1,2,3,4}$  is the torsion angle.

Additional disulfide bond engineering calculations were performed using the online server for Disulfide by Design version 2.11 (Dombkowski, 2003, *Bioinformatics*, 19:1852-1852, the entire contents of which are herein incorporated by reference.) Calculations were executed on both chains in the HjCel5A crystal structure (PDB ID 3QR3). Disulfide bonds principally contain four atoms linked in a linear fashion:  $C_\beta-S_\gamma-S_\gamma-C_\beta$ . In this calculation, a disulfide model is generated with fixed  $C_\beta-S_\gamma$  (1.81 Å) and  $S_\gamma-S_\gamma$  (2.04 Å) bond lengths and  $C_\beta-S_\gamma-S_\gamma$  (104.15°) bond angles. To initiate the calculation, a pair of residues is chosen. The  $\chi_3$  torsion angle, formed through rotating the  $C_\beta$  about the  $S_\gamma-S_\gamma$  bond, is allowed to vary until the  $C_\beta-C_\beta$  distance matches that observed in the crystal structure. Energies ( $E_{ij}$ ) are then calculated using the following equations:

$$E_{ij}=E(\chi_{1,i})+E(\chi_{1,j})+E(\theta_i)+E(\theta_j) \quad (1)$$

$$E(\chi_1)=1.4[1+\cos(3\chi_1)] \quad (2)$$

$$E(\chi_3)=4.0[1-\cos(2\chi_3+160)] \quad (3)$$

$$E(\theta)=55.0[\theta-\theta_0]^2 \quad (4)$$

Where i and j are residue positions,  $\theta$  is the  $C_\alpha-C_\beta-S_\gamma$  angle, and  $\theta_0=114.6^\circ$ . Energies are computed in kcal mol<sup>-1</sup> with

higher values corresponding to more favorable mutations. All calculations were performed with default settings.

Cel5A Plasmid Construction. Double mutants for disulfide engineering were constructed using a modified version of the Quickchange method in which two primer pairs are added to a single reaction. Additional details are as described in Example 1.

Thermostability/Activity Screen. Thermostability and activity screens were performed as described in Example 1.

Enzyme Purification. Cel5a proteins were cultured and purified as described in Example 1.

Park-Johnson Assay. The Park-Johnson assay was performed as described in Example 1.

T<sub>50</sub> Assay. The T<sub>50</sub> assay was performed as described in Example 1.

Single-Point Activity Assay. The single-point activity assay was performed as described in Example 1.

#### EXAMPLE 4

Identification of G189A, F191V, G239D, T233V, S242A, V265T, D271Y, S318P, and S322A, with Mutant Residue Numbering Based on SEQ ID NO: 5

Using a SCHEMA algorithm, a Cel5a chimera library was designed. *Hypocrea jecorina* Cel5a was recombined with three homologues from thermophilic fungi (*Phialophora* sp. G5, *Penicillium decumbens*, *Penicillium pinophilum*).

Out of this Cel5a chimera library, 25 chimeras were synthesized, appending to the N-terminus of each the cellulose binding module (CBM) from *H. jecorina* Cel5a, as well as a C-terminal His6 tag or purification. The chimeras were expressed in *Saccharomyces cerevisiae* and purified using column chromatography. It was observed that 23 out of the 25 chimeras were catalytically active, and the thermostabilities were measured by finding the temperature at which the enzymes loses half of its activity relative to that at its optimum temperature over a 2 hour (h) reaction period ("T<sub>450</sub>"). The T<sub>450</sub>'s for all of the functional chimeras was measured on crystalline cellulose (Avicel). The test set exhibits a range of thermostability values from which a regression model was built.

Regression modeling predicts a highly stable chimera. The revised linear regression model predicts that only two blocks are stabilizing relative to HjCel5A: blocks 6 and 7 from PgCel5A (+1.0° C. and +3.3° C., respectively). Combining these two blocks created chimera 110, which is 4.3° C. more stable and 18 mutations away from its closest parent, HjCel5A. Since chimera blocks may contain both stabilizing and destabilizing point mutations, each of the 18 single amino acid mutations were introduced into these blocks in HjCel5A individually and the thermostabilities of the mutant enzymes were tested.

As shown in FIG. 17, nine of the eighteen mutations were stabilizing, and seven were destabilizing, while two had no effect. The nine stabilizing HjCel5a point mutations include G189A, F191V, G239D, T233V, S242A, V265T, D271Y, S318P, and S322A, with mutant residue numbering based on SEQ ID NO: 5. Subsequently all of the stabilizing mutations (save for T233V, which compromised activity slightly) were combined to create chimera 110F. This chimera had a stability increase (as measured by T<sub>450</sub>) of 10.1° C. relative to HjCel5A (FIG. 18A). Its optimal temperature was also increased by ~10° C., and its activity was not compromised by thermostabilization.

To evaluate the improvement of 110F over industrially relevant time scales, its activity was compared to that of

HjCel5A over 60 h hydrolyses at 60° C. and 70° C. As shown in FIG. 18B, 110F displays more activity at both temperatures, and it remains highly active at 70° C. over a 60 h period, whereas wild-type HjCel5A is nearly inactive at 70° C.

Cel5A Plasmid Construction. Genes encoding *Hypocrea jecorina* Cel5A, *Phialophora* sp. G5 Cel5A, *Penicillium decumbens* Cel5A, and *Penicillium pinophilum* Cel5A were synthesized with *S. cerevisiae* codon bias (DNA 2.0, Menlo Park, Calif.), and cloned into the yeast secretion vector YEp352/PGK91-1- $\alpha$ ss as described previously by Komor et al. 2012, supra and Wu et al. 2013, *Biotech. and Bioeng.*, 110:1874-1883, the entire contents of both of which are herein incorporated by reference. Each gene had a C-terminal linker and carbohydrate binding module from *H. jecorina* Cel5A.

SCHEMA guided structure-based recombination. Gene sequences of *Hypocrea jecorina* Cel5A, *Phialophora* sp. G5 Cel5A, *Penicillium decumbens* Cel5A, and *Penicillium pinophilum* Cel5A were aligned using the MUSCLE multiple sequence alignment software, as described by Edgar, 2004, *Nucleic acids research*, 32:1792-1797, the entire contents of which are herein incorporated by reference. The structure of the catalytic domain of *H. jecorina* Cel5A (PDB structure 3QR3 chain A) (SEQ ID NO: 5) was used to build a map of amino acid contacts. (Lee et al., 2011, *Protein Science*, 20:1935-1940, the entire contents of which are herein incorporated by reference. A contact is defined as two amino acids having at least one non-hydrogen atom within 4.5 Å of each other. Libraries that minimized the average number of disrupted SCHEMA contacts in the resulting chimeras were designed using graph partitioning as described in Smith et al. 2013, *ACS Synthetic Biology*, Jun. 3, 2013, PMID: 23688124, the entire contents of which are herein incorporated by reference. A library design was chosen with an average SCHEMA energy (number of disrupted contacts) of 12.1 and an average of 55.4 amino acid mutations from the closest parent. The C-terminal linker and carbohydrate binding module from *H. jecorina* Cel5A was appended to each chimera.

Optimal Experimental Design. The Submodular Function Optimization Matlab toolbox (Ye et al., *J. Comput. Biol.*, 14: 777-790, the entire contents of which are herein incorporated by reference) was used to choose chimeras that had both low SCHEMA disruption and maximal mutual information between the sampled chimeras and the rest of the library, as described in Romero et al. 2012, *ACS Synthetic Biology*, 1:221-228, the entire contents of which are herein incorporated by reference.

Chimera Library Construction. Chimeras were constructed from 500 bp DNA fragments via overlap extension PCR, as described previously in Georgescu et al. 2003, Saturation mutagenesis: In *Directed evolution library creation*, pp. 75-83, Springer. The DNA fragments ("gBlocks") were synthesized by Integrated DNA Technologies (San Jose, Calif.). Codons were optimized for yeast expression using Gene Designer software from DNA 2.0 (Menlo Park, Calif.) (Vilalobos et al. 2006, *BMC Bioinformatics*, 7:285, the entire contents of which are herein incorporated by reference.) Genes were cloned into the YEp352/PGK91-1- $\alpha$ ss vector using Gibson assembly as described by Gibson 2011, *Methods Enzymology*, 498: 349-361, the entire contents of which are herein incorporated by reference.

Enzyme Purification. YEp352/PGK91-1- $\alpha$ ss vectors containing Cel5A chimeras were transformed into the BY4742  $\Delta$ kre2 strain of yeast (BY4742; Mat a; his3D1; leu2D0; lys2D0; ura3D0; YDR483w::kanMX4) obtained from EUROSCARF (Frankfurt, Germany). Yeast colonies expressing Cel5A with C-terminal His<sub>6</sub> tag were grown at 30° C.: first

overnight in 5 mL synthetic dextrose-uracil (SD-Ura) medium, then expanded into 50 mL SD-Ura (+50  $\mu$ g/mL kanamycin) medium for 24 h, and then expanded into 1 L YPD (+50  $\mu$ g/mL kanamycin) medium for an additional 48 h. Cultures were centrifuged at 4500 $\times$ g for 20 min and the supernatant was filtered with 0.2  $\mu$ m PES (polyethersulfone) filter unit from Nalgene (VWR, Radnor, Pa.). Protein was loaded onto 5 mL HisTrap columns and purified using an ÄKTAexpress chromatography system (GE Healthcare, Pittsburgh, Pa.). Purified cellulases were buffered-exchanged to 50 mM sodium acetate buffer pH 5.0 using Vivaspin 20 ultra-filtration spin tubes (GE Healthcare, Pittsburgh, Pa.). Protein concentrations were determined using A280, with theoretical extinction coefficients found using ProtParam on the ExPASy server as described in Gasteiger et al. 2005, *The Proteomics Protocols Handbook*, pp. 571-607, Springer, the entire contents of which are herein incorporated by reference.

$T_{A50}$  Thermostability Measurements. 100  $\mu$ L samples in 50 mM sodium acetate buffer, pH 5.0 containing 0.2  $\mu$ M Cel5A and 1% (w/v) Avicel were incubated at a range of temperatures for 2 h. A modified Park-Johnson reducing sugar assay was used to measure activity (Park and Johnson, 1949, *J. Biol. Chem.*, 181: 149-151, the entire contents of which are herein incorporated by reference.) In brief, reaction mixtures were spun at 1000 g for 5 min to remove Avicel. 50  $\mu$ L of supernatant was removed and transferred to a mixture of 100  $\mu$ L ferricyanide reagent (0.5 g/L K<sub>3</sub>Fe(CN)<sub>6</sub>, 34.84 g/L K<sub>2</sub>HPO<sub>4</sub>, pH 10.6) and 50  $\mu$ L carbonate-cyanide reagent (5.3 g/L Na<sub>2</sub>CO<sub>3</sub>, 0.65 g/L KCN). The reaction was heated at 95° C. for 15 min in an Eppendorf Mastercycler, and then cooled on ice for 5 min. 180  $\mu$ L of the reaction was removed and mixed with 90  $\mu$ L ferric iron solution (2.5 g/L FeCl<sub>3</sub>, 10 g/L polyvinyl pyrrolidone, 2 N H<sub>2</sub>SO<sub>4</sub>). After 2 min, absorbance at 595 nm was taken, using solutions of 0  $\mu$ M to 300  $\mu$ M cellobiose as standards.

$T_{A50}$  was determined by plotting activities against the temperature using Matlab (Mathworks, Natick, Mass., supra) and fitted using 4-parameter sigmoidal curves. The  $T_{A50}$  value is the temperature at which enzyme activity is halfway between optimal activity and no activity. Reported values were averaged from at least two independent measurements.

Cellulase Activity Measurements. All cellulase activity measurements were conducted in 50 mM sodium acetate buffer, pH 5.0. To determine activity-temperature profiles of Cel5A, samples containing 0.2  $\mu$ M of purified Cel5A and 1% (w/v) Avicel were incubated at 60 and 70° C. for 60 h. After hydrolysis, the reaction supernatants were sampled for reducing sugar concentrations via Nelson-Somogyi assay, using cellobiose as the reducing sugar standard as described in Somogyi, 1952, *J. Biol. Chem.*, 195: 19-23 and Nelson, 1944, *J. Biol. Chem.*, 153: 375-379, the entire contents of both of which are herein incorporated by reference. In brief: 50  $\mu$ L of reaction solution was added to 40  $\mu$ L carbonate-tartrate solution (144 g/L Na<sub>2</sub>SO<sub>4</sub>, 12 g/L potassium tartrate tetrahydrate, 24 g/L Na<sub>2</sub>CO<sub>3</sub>, 16 g/L NaHCO<sub>3</sub>) and 10  $\mu$ L copper solution (180 g/L Na<sub>2</sub>SO<sub>4</sub>, 20 g/L CuSO<sub>4</sub>·5H<sub>2</sub>O) and heated to 95° C. for 15 min in an Eppendorf Mastercycler. The reaction was placed on ice for 5 min and then mixed with 50  $\mu$ L arsenomolybdate solution (50 g/L (NH<sub>4</sub>)<sub>2</sub>MoO<sub>4</sub>, 1.5 N H<sub>2</sub>SO<sub>4</sub>, 6 g/L NaH<sub>2</sub>AsO<sub>4</sub>). After mixing, absorbance at 520 nm was read, using 0 to 2 mM cellobiose solutions as standards.

## EXAMPLE 5

### Combination Cel5a Mutants

First Generation Combination Cel5a endoglucanase Mutants. Using the information from Examples 1-4, a series

of combination constructs were synthesized containing highly stabilizing mutations. Mutations demonstrating a  $\Delta T_{50} \geq 0.5^\circ \text{C}$ . were selected for incorporation. If several stabilizing mutations appeared in the same region within the H<sub>j</sub>Cel5A structure, the mutations with the highest  $\Delta T_{50}$  values were generally retained. In ambiguous situations where interact on between two sites could not clearly be ascertained, several alternative constructs were tested. A chosen set included 13 possible mutations: T57N, N76P, T80E, S139P, N155E, G189S, K219Q, G239N, D271F, Y278F, G293A, S309F, and S318P, with mutant residue numbering based on SEQ ID NO: 5.

These mutations were incrementally added to the WT sequence with the least and most mutated constructs containing 1 and 13 mutations, respectively (Table V), with mutant residue numbering based on SEQ ID NO: 5. Following cloning, mutants were tested for activity with the same screen employed to detect point mutations (FIG. 19A). All of the constructs showed improvements in activity over WT.

G189S, D271F, Y278F, and G293A) (with mutant residue numbering based on SEQ ID NO: 5), construct 16 demonstrated the greatest performance on the activity screen. Construct 20 contains all 13 possible mutations and was projected to demonstrate the greatest thermostability. These two combination mutants were expressed, purified and tested to assess their hydrolytic capabilities on Avicel over a gradient of temperatures (FIG. 19B), obtain their  $T_{50}$  values (FIG. 19C), and determine their activity at  $60^\circ \text{C}$ . after 2 hours (FIG. 19D). Table VI provides a summary of these results. The 9-point mutant (construct 16 of Table V) shows a  $10^\circ \text{C}$ . increase in both the optimal operating temperature ( $T_{opt}$ ) and in  $T_{50}$  relative to WT. As expected, the 13-point mutant demonstrates even greater improvement in thermostability with a  $16^\circ \text{C}$ . increase in  $T_{opt}$  and a  $14^\circ \text{C}$ . increase in  $T_{50}$  relative to WT. These increases, however, were accompanied by decreases in activity of 4% for the 9-point mutant and 17% for the 13-point mutant. Analysis revealed that the 13-point mutant contains more mutations that are known from previous experiments to decrease activity than the 9-point mutant.

TABLE V

Composition of Combination Constructs														
Construct	# of Mutations	T57N	N76P	T80E	S139P	N155E	G189S	K219Q	G239N	D271F	Y278F	G293A	S309F	S318P
1	1		x											
2	1				x									
3	1					x								
4	2		x		x									
5	4	x	x		x	x								
6	5	x	x		x	x	x							
7	5	x	x		x	x		x						
8	5	x	x		x	x			x					
9	5	x	x		x	x				x				
10	6	x	x		x	x				x	x			
11	6	x	x		x	x	x	x						
12	6	x	x		x	x	x		x					
13	6	x	x		x	x	x	x	x					
14	7	x	x		x	x	x			x	x			
15	8	x	x		x	x	x			x	x	x		
16	9	x	x	x	x	x	x			x	x	x		
17	10	x	x		x	x	x	x	x	x	x	x		
18	11	x	x	x	x	x	x	x	x	x	x	x		
19	12	x	x	x	x	x	x	x	x	x	x	x		x
20	13	x	x	x	x	x	x	x	x	x	x	x	x	x

Construct	# of Mutations	T57N	T80E	N155E	G189S	G239E	D271Y	G293A	S309L/W	S318E/P	S79P	V101I	S133R	E53D
s13pt1	13	x	x	x	x	x	x	x	x(L)	x(P)	x	x	x	x
s13pt2	13	x	x	x	x	x	x	x	x(W)	x(P)	x	x	x	x
s13pt3	13	x	x	x	x	x	x	x	x(L)	x(E)	x	x	x	x
s13pt4	13	x	x	x	x	x	x	x	x(W)	x(E)	x	x	x	x

Two constructs predicted to have either high activity or thermostability were chosen for further characterization. With nine mutations (T57N, N76P, T80E, S139P, N155E,

As such, the drop in activity likely originates from the collective effects of point mutations rather than incompatibilities between the mutations.

TABLE VI

Characterization of Combination Mutants							
Construct	# of Mutations	$T_{50,mut}$ ( $^\circ \text{C}$ .)	$\Delta T_{50}$ ( $^\circ \text{C}$ .)	$T_{opt}$ ( $^\circ \text{C}$ .)	Activity ( $\mu\text{M}$ Cellobiose Equivalents)	$\Delta$ Activity ( $\mu\text{M}$ Cellobiose Equivalents)	Expression Level Relative to WT
WT	0	$69.6 \pm 1.0$	—	64	$216.7 \pm 10.6$	—	—
9 pt	9	$81.5 \pm 0.4$	$10.0 \pm 0.8$	74	$184.2 \pm 5.7$	-9.6	3.5
13 pt	13	$86.8 \pm 0.6$	$13.8 \pm 0.9$	80	$157.4 \pm 1.4$	-36.3	4.1
20 pt	20	$90.0 \pm 0.1$	$16.8 \pm 0.7$	82	$146.6 \pm 9.8$	-47.1	2.7
s13pt1	13	$85.3 \pm 0.2$	$14.9 \pm 0.7$	78	$192.7 \pm 6.9$	-1.0	4.1

TABLE VI-continued

Characterization of Combination Mutants							
Construct	# of Mutations	T <sub>50,mut</sub> (° C.)	ΔT <sub>50</sub> (° C.)	T <sub>opt</sub> (° C.)	Activity (μM Cellobiose Equivalents)	ΔActivity (μM Cellobiose Equivalents)	Expression Level Relative to WT
s13pt2	13	85.6 ± 0.4	15.4 ± 0.7	78	221.1 ± 5.9	27.4	4.1
s13pt3	13	83.0 ± 0.4	12.0 ± 0.8	75	231.2 ± 5.6	37.5	5.9
s13pt4	13	83.9 ± 0.2	12.2 ± 0.7	75	262.6 ± 11.5	68.8	3.7

To ascertain whether additional mutagenesis could counter these activity decreases, a 20-point mutant was constructed containing mutations with  $\Delta T_{50} \geq 0^\circ \text{C}$ . This combination mutant contains all 13 of the previously incorporated mutations plus T18P, G64A, S79P, V101I, S133R, D13E, and E53D. Individually, many of the less stabilizing mutations show improvements in activity; summing the changes in activity measured for these point mutations gives a net activity increase of 20.7 μM cellobiose equivalents. As such, that possibility that these less stabilizing mutations might boost thermostability while rescuing activity was tested. While the 20-point mutant shows a  $\Delta T_{50}$  of 16.8° C. (FIG. 19C), its activity at 60° C. decreases even further to 47.1 μM cellobiose equivalents below WT (22% of WT activity) (FIG. 19D). This result suggests that activity losses from individual mutations permanently accumulate and cannot be rescued by adding mutations that improve activity in isolation.

Although these initial combination mutants exhibit diminished activity, it was hypothesized that their enhanced thermostability might improve performance in longer assays. Due to its relatively modest decrease in activity, the 9-point mutant (construct 16 of Table V) was chosen for 60-hour activity tests on Avicel at 60 and 70° C. (FIGS. 19E, 19F). This construct demonstrates a nine-fold activity improvement over WT at the elevated temperature. When compared to WT hydrolysis at 60° C., the total product yield improves by about 134 μM of cellobiose equivalents (24% increase, 1.2 fold improvement).

Second Generation Combination Cel5a endoglucanase Mutants. Using the 9-point mutant as a template, a second-generation of combination mutants was constructed by excluding all mutations detrimentally affecting activity. The process involved reverting mutations N76P, S139P, K219Q, G239N, D271F, Y278F, and S309F back to the WT residue or, if available, a less stable, more active alternate. Mutations appearing in the four final second generation 13-point combination mutants (s13pt 1-4) are summarized in Table V, with mutant residue numbering based on SEQ ID NO: 5.

The second-generation combination mutants performed as well or better than WT on all tested metrics. All four constructs show enhancements in thermostability with s13pt2 demonstrating the highest increase ( $\Delta T_{50}=15.4^\circ \text{C}$ ) (Table VI, FIGS. 20A, 20B, 20C, 20D). Activity at 60° C. improved over WT for all constructs except s13pt1 (FIG. 20C). In this case, activity declined slightly by 1.0 μM of cellobiose. The combination mutants exhibit dramatic improvements in  $T_{opt}$ . In a two-hour activity test, s13pt1/2 and s13pt3/4 optimally function at 78 ( $\neq T_{opt}=14^\circ \text{C}$ ) and 75 degrees ( $\Delta T_{opt}=11^\circ \text{C}$ ), respectively. Additionally, all of the mutants show ~4-6 fold increases in expression level over WT.

To approximate industrial reactions, 60-hour hydrolysis experiments were conducted on the constructs with the highest activity (s13pt4) and thermostability (s13pt2). At 60° C., both combination mutants performed similarly to WT (FIG. 21A). This improvement dramatically increased at 70° C.

(FIG. 21B). Compared to the WT performance at the same temperature, s13pt4 and s13pt2 exhibit ~9.5-10 fold increases in activity. Moreover, the combination mutants improve yield by 358-414 μM cellobiose equivalents (~60% increase) over the maximal amount possible using the WT enzyme.

As shown, these combination mutants optimally perform at 75-78° C. in the 2-hour hydrolysis experiments. A long-term hydrolysis was tested to see if these mutant enzymes improved at higher temperatures. However, after 60-hours of hydrolysis, less activity was observed at 75 and 78° C. than at 70° C. (FIGS. 21C, 21D).

**Cel5A Plasmid Construction.** Cel5a mutants were constructed as described in Example 1.

**Thermostability/Activity Screen.** Thermostability and activity screens were performed as described in Example 1.

The Park-Johnson Assay was performed as described in Example 1.

**Enzyme Purification.** Cel5a proteins were cultured and purified as described in Example 1.

$T_{50}$  Assay was performed as described in Example 1.

$T_{opt}$  Assay. Determination of the temperature yielding the maximum activity proceeded through incubating enzyme with Avicel for two hours at a gradient of temperatures, then determining sugar release with the Park-Johnson assay. In a 96-well PCR plate, 40 μL of purified enzyme were combined with 60 μL of 1.5% Avicel suspended in cellulase buffer (100 mL 50 mM sodium acetate, pH 5.0). Samples were incubated for 2 hours across a 20° C. gradient centered around a temperature projected to capture the peak of activity based on  $T_{50}$  values. Activity was assessed from 25 μL of supernatant using the Park-Johnson assay.

Single-Point Activity Assay was performed as described in Example 1.

**60-Hour Activity Assay.** To assess activity over a constant temperature for 60 hours, enzyme and substrate mixtures were combined in individual PCR tubes and frozen to arrest hydrolysis. In each tube, 40 μL of purified enzyme at a concentration of 0.5 μM was combined with 60 μL of 1.67% Avicel suspended in cellulase buffer (100 mL 50 mM sodium acetate, pH 5.0). Incubation occurred in a PCR block pre-heated before adding samples to prevent background activity. Time points were collected at 0, 4, 8, 16, 24, 36, 48, and 60 hours. Following hydrolysis, the reactions were thawed and moved to a 96-well plate to facilitate centrifugation. Supernatants were robotically collected. Cellobiose standards containing 0.0, 166.6, 333.3, 500.0, 833.3, 1000.0, 1500.0, and 2000.0 μM of cellobiose and 50 μL of the reaction supernatants were assessed for reducing sugar concentrations via the Nelson-Somogyi assay as described in Smogyi, 1952, *J. Biol. Chem.*, 195: 19-23 and Nelson, 1944, *J. Biol. Chem.*, 153: 375-379, the entire contents of both of which are herein incorporated by reference. All experiments were performed in triplicate.

## A Thermostable Cel5 Endoglucanase

Out of the point mutations identified in Examples 1-4, sixteen thermostabilizing mutations that did not compromise activity measured at 60° C. were combined into a single variant. If two suitable mutations occupied the same site, the more thermostabilizing of the two was selected. This combination HjCel5A mutant was termed: OptCel5a, which includes: T57N, E53D, S79P, T80E, V101I, S133R, N155E, G189S, F191V, T233V, G239E, V265T, D271Y, G293A, S309W, and S318P point mutations, with mutant residue numbering based on SEQ ID NO: 5.

OptCel5a was transformed into the BY4742 Δkre2 strain of *Saccharomyces cerevisiae*, a glycosylation-deficient strain that secretes proteins that have similar glycosylation patterns to *H. jecorina* enzymes (Heinzelman et al. 2009, *PNAS*, 106: 5610-5615, the entire contents of which are herein incorporated by reference). When expressed and purified, OptCel5A has an optimal temperature of 81.1° C. when used to hydrolyze crystalline cellulose (Avicel) for 2 h (FIG. 22A). This makes OptCel5A more than 17° C. more thermostable than wild type HjCel5A, more than 7° C. more stable than the 110F chimera 110F (Example 4, FIG. 18A), and 6° C. more stable than s13pt4 (Example 5, FIGS. 20C, 20D).

The activity of OptCel5A and wild type HjCel5A was assayed over a 60 h hydrolysis at both 60° C. and 70° C. OptCel5A had highest activity at 70° C., hydrolyzing more than 1.5 times as much cellulose as HjCel5A at its optimal temperature of 60° C. (FIG. 22B). OptCel5A is therefore compatible with the previously engineered thermostable Cel6A and Cel7A, which both operate optimally at 70° C. in 60 h hydrolysis experiments (Wu and Arnold, 2013, *Biotech. and Bioeng.*, 110:1874-1883, the entire contents of which are herein incorporated by reference).

Plasmids and Strains. The gene encoding wild type Cel5A (including its cellulose binding module) was synthesized with *S. cerevisiae* codon optimization (DNA 2.0). Plasmids were transformed into the BY4742 Δkre2 strain of yeast (BY4742; Mat a; his3D1; leu2D0; lys2D0; ura3D0; YDR483w::kanMX4) obtained from EUROSCARF.

Enzyme Purification. Yeast colonies expressing Cel5A with C-terminal His<sub>6</sub> tag and was grown at 30° C.: first overnight in 5 mL SD-Ura medium, expanded into 50 mL SD-Ura (+50 μg/mL kanamycin) medium for 24 h, and then expanded into 1 L YPD (+50 μg/mL kanamycin) medium for an additional 48 h. Cultures were centrifuged at 4500 g for 20 min, and the supernatant was filtered with 0.2 mm PES filter unit from Nalgene. Protein was purified using 5 mL HisTrap columns (GE Healthcare). Purified cellulases were buffered-exchanged to 50 mM sodium acetate buffer pH 5.0 using Vivaspin 20 ultrafiltration spin tubes (GE Healthcare). Protein concentrations were determined using A<sub>280</sub>, with extinction coefficients calculated using ProtParam on the ExPASy server (Gasteiger et al. 2005, *supra*).

Thermostability Measurements. 100 μL samples in 50 mM sodium acetate buffer, pH 5.0 containing 0.2 μM Cel5A and 1% (w/v) Avicel were incubated at a range of temperatures for 2 h in an Eppendorf Mastercycler. A modified Park-Johnson reducing sugar assay was used to measure activity (Park and Johnson, 1949, *JBC*, 181:149-151, the entire contents of which are herein incorporated by reference). Briefly, reaction mixtures were spun at 1000 g for 5 min to remove Avicel. 50 μL of supernatant was removed and transferred to a mixture of 100 μL ferricyanide reagent (0.5 g/L K<sub>3</sub>Fe(CN)<sub>6</sub>, 34.84 g/L K<sub>2</sub>HPO<sub>4</sub>, pH 10.6) and 50 μL carbonate-cyanide reagent (5.3

g/L Na<sub>2</sub>CO<sub>3</sub>, 0.65 g/L KCN). The reaction was heated at 95° C. for 15 min in an Eppendorf Mastercycler, and then cooled on ice for 5 min. 180 μL of the reaction was removed and mixed with 90 μL ferric iron solution (2.5 g/L FeCl<sub>3</sub>, 10 g/L polyvinyl pyrrolidone, 2 N H<sub>2</sub>SO<sub>4</sub>). After 2 min, absorbance at 595 nm was taken, using solutions of 0 μM to 300 μM cellobiose as standards.

Cellulose Activity Measurements. All cellulase activity measurements were conducted in 50 mM sodium acetate buffer, pH 5.0. Constant temperature was maintained using an Eppendorf Mastercycler. To determine activity-temperature profiles of Cel5A, samples containing 0.2 μM of purified Cel5a and 1% (w/v) Avicel were incubated at 60 and 70 20 C. for 60 h. After hydrolysis, reaction supernatants were sampled for reducing sugar concentrations via a modified Nelson-Somogyi assay (Green et al. 1989, *Anal. Biochem.*, 182: 197-199, the entire contents of which are herein incorporated by reference). Briefly, 50 μL of reaction solution was added to 40 μL carbonate-tartrate solution (144 g/L Na<sub>2</sub>SO<sub>4</sub>, 12 g/L potassium tartrate tetrahydrate, 24 g/L Na<sub>2</sub>CO<sub>3</sub>, 16 g/L NaHCO<sub>3</sub>) and 10 μL copper solution (180 g/L Na<sub>2</sub>SO<sub>4</sub>, 20 g/L CuSO<sub>4</sub>·5H<sub>2</sub>O) and heated to 95° C. for 15 min in an Eppendorf Mastercycler. The reaction was placed on ice for 5 min and then mixed with 50 μL arsenomolybdate solution (50 g/L (NH<sub>4</sub>)<sub>2</sub>MoO<sub>4</sub>, 1.5 N H<sub>2</sub>SO<sub>4</sub>, 6 g/L NaH<sub>2</sub>AsO<sub>4</sub>). After mixing, absorbance at 520 nm was read, using 0 to 2 mM cellobiose solutions as standards.

Cellulose hydrolysis to determine the activity of optimized engineered and wild type cellulase mixtures was carried out on 1%, 3%, and 5% Avicel at 60° C. and 70° C. Samples were taken at 0 h, 4 h, 8 h, 15 h, 24 h, 36 h, 48 h, and 60 h, and reducing sugar concentration was quantified as described above.

Data analysis. Cellulase activity and thermostability data were plotted using Microsoft Excel. Synergy plots were made in Matlab (The Mathworks, Inc.), using the Templot package developed by Carl Sandrock.

As disclosed throughout as evidenced by, for example, the cellulase activity of a variant Cel5a shown in FIGS. 22A, 22B, a variant Cel5a enzyme has improved thermostability and activity compared to wild type.

While the present invention has been illustrated and described with reference to certain exemplary embodiments, those of ordinary skill in the art will understand that various modifications and changes may be made to the described embodiments without departing from the spirit and scope of the present invention, as defined in the following claims.

## SEQUENCE LISTING

## SEQ ID NO: 1

HjCel5a Nucleotide Sequence GGCGTTAFATTTFCCG-  
GTGTTAATATTGCTGGTTTTGACTTCG-  
GTTGCACTACCGATGGCACTTGTGTTACT TCTAAG-  
GTCTATCCTCCGCTTAAGAAGTTTACGGGTTCCAAC  
AAGTATCCTGATGATAGGGCAGATGCAACA  
TTTTGTTAAIGAAGACGGTATGA-  
CAATATTTTCGTTTGCCCTGTTGGATG-  
GCAATATCTGGTCAACAATAACCT GGGAGGTAATT-  
TAGATAGTACCTCTATCTCCAAATATACAATTGGTCC  
AAGGTTGTCTATCCTTAGGTGCAT ATTGTATTGTC-  
GATATACATAATTATGCTAGATGGAAATG-  
GCGGTATTATTGGTCAAGGCGGTCCAACAAACG  
CGCAATTTACTTCATTGTGAGCAGTTG-  
GCTAGTAAATACGCGTCACAGTC-  
CAGGGTTTGGTTTGGGAATTATG AATGAGCCACAC-

33

GATGTTAACATTAATACCTGGGCTGCTACCGTTCAA  
 GAAGTTGTACAGCAATTAGAAAT GTGGGC-  
 TACGTCCCAGTTTATCAGTCTACCTGGTAATGATTG  
 GCAATCTGCTGGTGTCTTTCATTCTGACGG CAGT-  
 GCCGCTGCGTTGTGCGCAAGTAACTAATC- 5  
 CAGATGGCTCCACAAC-  
 TAATCTAATTTTCGACGTGCATAA  
 GTATTTGGATTTCGATAATAGTGGTACT-  
 CATGCAGAGTGTACTACTAACAATATC-  
 GATGGTGCCTAGCCCGT TGGCAACCTGGTTACGT- 10  
 CAAAACAATAGACAAGCAATATTGACGGAAACCGG  
 TGGTGGTAATGTACAAAGTT GTATTACAGGATATGT-  
 GTCAACAAAATACAGTACCTTAAC-  
 CAAAACCTCAAGTTTACTTAGGCTACGTTGGCTGG

34

GGTGCTGGTTCCTTCGACAGTACT-  
 TACGTTTTGACTGAGACACCTACATC-  
 CAGTGGTAATAGTTGGACCGAT ACTTCTTTGG-  
 TATCTTCTTGCTTAGCTAGAA  
 SEQ ID NO: 2  
 AR GGCTAGACAACAAACAGTATGGGGT-  
 CAATGTGGTGGTATTGGATGGTCTGGTCCGA-  
 CAAACTGTGCTCCAGGCTCGGCATGTTT- 10  
 GACACTAAATCCATATTACGCTCAATGTATCCCTGG  
 CGCTACCACTATAACAACTTCTACTA-  
 GACCACCTTCTGGTCCGACGACAACCTA-  
 CAAGGGCTACCTCAA CCTCTTCCTCTACAC-  
 CCCCTACTTCCAGC

## SEQUENCE LISTING

&lt;160&gt; NUMBER OF SEQ ID NOS: 5

&lt;210&gt; SEQ ID NO 1

&lt;211&gt; LENGTH: 967

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Hypocrea jecorina

&lt;400&gt; SEQUENCE: 1

```

ggcgtagat ttgcgggtgt taatattgct ggttttgact tcggttgca caccgatggc      60
acttgtgtta cttctaaggt ctatcctccg cttaagaact ttacgggttc caacaactat      120
cctgatgata gggcagatgc aacattttgt taatgaagac ggtatgacaa tatttcgttt      180
gctgttga tggaatatc tggtaacaa taacctggga ggtaatttag atagtacctc      240
tatctccaaa tatacaattg gtccaaggtt gtctatcctt aggtgcata tgtattgtcg      300
atacacataa ttatgctaga tggaatggcg gtattattgg tcaaggcggt ccaacaaacg      360
cgcaatttac ttcattgtga gcagttggct agtaaatacg cgtcacagtc cagggtttgg      420
tttgaatta tgaatgagcc acacgatgtt aacattaata cctgggctgc taccgttcaa      480
gaagttgtca cagcaattag aaatgtgggc tacgtccag tttatcagtc tacctggtaa      540
tgattggcaa tctgtgtgtg ctttcatttc tgacggcagt gccgctgcgt tgcgcaagt      600
aactaatcca gatggctcca caactaatct aattttcgac gtgcataagt atttgattc      660
ggataatagt ggtactcatg cagagtgtac tactaacaat atcgatggtg cctagcccgt      720
tggaacctg gttactgtca aacaatagac aagcaatatt gacggaaacc ggtggtggtg      780
atgtacaaag ttgtattcag gatattgtgc aacaaataca gtaccttaac caaaactcaa      840
gtttacttag gtaactgttg ctgggggtgct gggtccttcg acagtactta cgttttgact      900
gagacaccta catccagtgg taatagttgg accgatactt ctttggtatc ttcttgctta      960
gctagaa                                           967

```

&lt;210&gt; SEQ ID NO 2

&lt;211&gt; LENGTH: 217

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Added sequence to Cel5a

&lt;400&gt; SEQUENCE: 2

```

ggctagacaa caaacagtat ggggtcaatg tgggtgtatt ggtggtctg gtccgacaaa      60
ctgtgctcca ggctcgcat gttcgacact aaatccatat tacgtcaat gtatccctgg      120
cgctaccact ataacaactt ctactagacc accttctggt ccgacgacaa ctacaagggc      180

```



-continued

---

tacctcaacc tcttctcteta caccctctac ttccagc

217

&lt;210&gt; SEQ ID NO 3

&lt;211&gt; LENGTH: 418

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Hypocrea jecorina

&lt;400&gt; SEQUENCE: 3

Met Asn Lys Ser Val Ala Pro Leu Leu Leu Ala Ala Ser Ile Leu Tyr  
1 5 10 15

Gly Gly Ala Ala Ala Gln Gln Thr Val Trp Gly Gln Cys Gly Gly Ile  
20 25 30

Gly Trp Ser Gly Pro Thr Asn Cys Ala Pro Gly Ser Ala Cys Ser Thr  
35 40 45

Leu Asn Pro Tyr Tyr Ala Gln Cys Ile Pro Gly Ala Thr Thr Ile Thr  
50 55 60

Thr Ser Thr Arg Pro Pro Ser Gly Pro Thr Thr Thr Thr Arg Ala Thr  
65 70 75 80

Ser Thr Ser Ser Ser Thr Pro Pro Thr Ser Ser Gly Val Arg Phe Ala  
85 90 95

Gly Val Asn Ile Ala Gly Phe Asp Phe Gly Cys Thr Thr Asp Gly Thr  
100 105 110

Cys Val Thr Ser Lys Val Tyr Pro Pro Leu Lys Asn Phe Thr Gly Ser  
115 120 125

Asn Asn Tyr Pro Asp Gly Ile Gly Gln Met Gln His Phe Val Asn Asp  
130 135 140

Asp Gly Met Thr Ile Phe Arg Leu Pro Val Gly Trp Gln Tyr Leu Val  
145 150 155 160

Asn Asn Asn Leu Gly Gly Asn Leu Asp Ser Thr Ser Ile Ser Lys Tyr  
165 170 175

Asp Gln Leu Val Gln Gly Cys Leu Ser Leu Gly Ala Tyr Cys Ile Val  
180 185 190

Asp Ile His Asn Tyr Ala Arg Trp Asn Gly Gly Ile Ile Gly Gln Gly  
195 200 205

Gly Pro Thr Asn Ala Gln Phe Thr Ser Leu Trp Ser Gln Leu Ala Ser  
210 215 220

Lys Tyr Ala Ser Gln Ser Arg Val Trp Phe Gly Ile Met Asn Glu Pro  
225 230 235 240

His Asp Val Asn Ile Asn Thr Trp Ala Ala Thr Val Gln Glu Val Val  
245 250 255

Thr Ala Ile Arg Asn Ala Gly Ala Thr Ser Gln Phe Ile Ser Leu Pro  
260 265 270

Gly Asn Asp Trp Gln Ser Ala Gly Ala Phe Ile Ser Asp Gly Ser Ala  
275 280 285

Ala Ala Leu Ser Gln Val Thr Asn Pro Asp Gly Ser Thr Thr Asn Leu  
290 295 300

Ile Phe Asp Val His Lys Tyr Leu Asp Ser Asp Asn Ser Gly Thr His  
305 310 315 320

Ala Glu Cys Thr Thr Asn Asn Ile Asp Gly Ala Phe Ser Pro Leu Ala  
325 330 335

Thr Trp Leu Arg Gln Asn Asn Arg Gln Ala Ile Leu Thr Glu Thr Gly  
340 345 350

Gly Gly Asn Val Gln Ser Cys Ile Gln Asp Met Cys Gln Gln Ile Gln  
355 360 365

-continued

---

Tyr Leu Asn Gln Asn Ser Asp Val Tyr Leu Gly Tyr Val Gly Trp Gly  
 370 375 380  
 Ala Gly Ser Phe Asp Ser Thr Tyr Val Leu Thr Glu Thr Pro Thr Gly  
 385 390 395 400  
 Ser Gly Asn Ser Trp Thr Asp Thr Ser Leu Val Ser Ser Cys Leu Ala  
 405 410 415  
 Arg Lys  
 <210> SEQ ID NO 4  
 <211> LENGTH: 340  
 <212> TYPE: PRT  
 <213> ORGANISM: Hypocrea jecorina  
 <400> SEQUENCE: 4  
 Met Gly Val Arg Phe Ala Gly Val Asn Ile Ala Gly Phe Asp Phe Gly  
 1 5 10 15  
 Cys Thr Thr Asp Gly Thr Cys Val Thr Ser Lys Val Tyr Pro Pro Leu  
 20 25 30  
 Lys Asn Phe Thr Gly Ser Asn Asn Tyr Pro Asp Gly Ile Gly Gln Met  
 35 40 45  
 Gln His Phe Val Asn Glu Asp Gly Met Thr Ile Phe Arg Leu Pro Val  
 50 55 60  
 Gly Trp Gln Tyr Leu Val Asn Asn Asn Leu Gly Gly Asn Leu Asp Ser  
 65 70 75 80  
 Thr Ser Ile Ser Lys Tyr Asp Gln Leu Val Gln Gly Cys Leu Ser Leu  
 85 90 95  
 Gly Ala Tyr Cys Ile Val Asp Ile His Asn Tyr Ala Arg Trp Asn Gly  
 100 105 110  
 Gly Ile Ile Gly Gln Gly Gly Pro Thr Asn Ala Gln Phe Thr Ser Leu  
 115 120 125  
 Trp Ser Gln Leu Ala Ser Lys Tyr Ala Ser Gln Ser Arg Val Trp Phe  
 130 135 140  
 Gly Ile Met Asn Glu Pro His Asp Val Asn Ile Asn Thr Trp Ala Ala  
 145 150 155 160  
 Thr Val Gln Glu Val Val Thr Ala Ile Arg Asn Ala Gly Ala Thr Ser  
 165 170 175  
 Gln Phe Ile Ser Leu Pro Gly Asn Asp Trp Gln Ser Ala Gly Ala Phe  
 180 185 190  
 Ile Ser Asp Gly Ser Ala Ala Ala Leu Ser Gln Val Thr Asn Pro Asp  
 195 200 205  
 Gly Ser Thr Thr Asn Leu Ile Phe Asp Val His Lys Tyr Leu Asp Ser  
 210 215 220  
 Asp Asn Ser Gly Thr His Ala Glu Cys Thr Thr Asn Asn Ile Asp Gly  
 225 230 235 240  
 Ala Phe Ser Pro Leu Ala Thr Trp Leu Arg Gln Asn Asn Arg Gln Ala  
 245 250 255  
 Ile Leu Thr Glu Thr Gly Gly Gly Asn Val Gln Ser Cys Ile Gln Asp  
 260 265 270  
 Met Cys Gln Gln Ile Gln Tyr Leu Asn Gln Asn Ser Asp Val Tyr Leu  
 275 280 285  
 Gly Tyr Val Gly Trp Gly Ala Gly Ser Phe Asp Ser Thr Tyr Val Leu  
 290 295 300  
 Thr Glu Thr Pro Thr Ser Ser Gly Asn Ser Trp Thr Asp Thr Ser Leu  
 305 310 315 320

-continued

---

Val Ser Ser Cys Leu Ala Arg Lys Gly Gly Ser Gly Ser Gly His His  
                   325                  330                  335

His His His His  
                   340

<210> SEQ ID NO 5  
 <211> LENGTH: 339  
 <212> TYPE: PRT  
 <213> ORGANISM: *Hypocrea jecorina*

<400> SEQUENCE: 5

Gly Val Arg Phe Ala Gly Val Asn Ile Ala Gly Phe Asp Phe Gly Cys  
   1                  5                  10                  15

Thr Thr Asp Gly Thr Cys Val Thr Ser Lys Val Tyr Pro Pro Leu Lys  
           20                  25                  30

Asn Phe Thr Gly Ser Asn Asn Tyr Pro Asp Gly Ile Gly Gln Met Gln  
           35                  40                  45

His Phe Val Asn Glu Asp Gly Met Thr Ile Phe Arg Leu Pro Val Gly  
           50                  55                  60

Trp Gln Tyr Leu Val Asn Asn Asn Leu Gly Gly Asn Leu Asp Ser Thr  
   65                  70                  75                  80

Ser Ile Ser Lys Tyr Asp Gln Leu Val Gln Gly Cys Leu Ser Leu Gly  
                   85                  90                  95

Ala Tyr Cys Ile Val Asp Ile His Asn Tyr Ala Arg Trp Asn Gly Gly  
           100                  105                  110

Ile Ile Gly Gln Gly Gly Pro Thr Asn Ala Gln Phe Thr Ser Leu Trp  
           115                  120                  125

Ser Gln Leu Ala Ser Lys Tyr Ala Ser Gln Ser Arg Val Trp Phe Gly  
           130                  135                  140

Ile Met Asn Glu Pro His Asp Val Asn Ile Asn Thr Trp Ala Ala Thr  
   145                  150                  155                  160

Val Gln Glu Val Val Thr Ala Ile Arg Asn Ala Gly Ala Thr Ser Gln  
                   165                  170                  175

Phe Ile Ser Leu Pro Gly Asn Asp Trp Gln Ser Ala Gly Ala Phe Ile  
           180                  185                  190

Ser Asp Gly Ser Ala Ala Ala Leu Ser Gln Val Thr Asn Pro Asp Gly  
           195                  200                  205

Ser Thr Thr Asn Leu Ile Phe Asp Val His Lys Tyr Leu Asp Ser Asp  
           210                  215                  220

Asn Ser Gly Thr His Ala Glu Cys Thr Thr Asn Asn Ile Asp Gly Ala  
   225                  230                  235                  240

Phe Ser Pro Leu Ala Thr Trp Leu Arg Gln Asn Asn Arg Gln Ala Ile  
                   245                  250                  255

Leu Thr Glu Thr Gly Gly Gly Asn Val Gln Ser Cys Ile Gln Asp Met  
           260                  265                  270

Cys Gln Gln Ile Gln Tyr Leu Asn Gln Asn Ser Asp Val Tyr Leu Gly  
           275                  280                  285

Tyr Val Gly Trp Gly Ala Gly Ser Phe Asp Ser Thr Tyr Val Leu Thr  
           290                  295                  300

Glu Thr Pro Thr Ser Ser Gly Asn Ser Trp Thr Asp Thr Ser Leu Val  
   305                  310                  315                  320

Ser Ser Cys Leu Ala Arg Lys Gly Gly Ser Gly Ser Gly His His His  
                   325                  330                  335

His His His

---

What is claimed is:

1. A composition comprising a variant Cel5a endoglucanase having a temperature at which half of the maximal protein activity remains ( $T_{50}$ ) higher than a  $T_{50}$  of wild type Cel5a endoglucanase of SEQ ID NO: 3 and having increased enzymatic activity relative to enzymatic activity of the wild type Cel5a endoglucanase of SEQ ID NO: 3, the variant Cel5a endoglucanase comprising point mutations in a wild type Cel5a amino acid sequence of SEQ ID NO: 4 selected from the group consisting of

(a) G293A and S318P/Q/E and

(b) G293A and S318P/Q/E and one or more residues selected from the group consisting of E53D, T57N, S79P, T80E, V101I, S133R, N153D, N155E, T156E, G189S, F191V, T233V, G239E, V265T, D271Y, and S309W/L, in which the amino acid residue numbering is based on the numbering of SEQ ID NO: 5.

2. The composition of claim 1, wherein the variant Cel5a endoglucanase has a  $T_{50}$  above 70° C.

3. The composition of claim 1, wherein the variant Cel5a endoglucanase has a  $T_{50}$  of about 1° C. to about 17° C. above the  $T_{50}$  of the wild type Cel5a.

4. The composition of claim 1, wherein the increased enzymatic activity is a net activity increase in amount of digested cellulose relative to the wild type enzymatic activity of Cel5a endoglucanase of SEQ ID NO: 3.

5. The composition of claim 4, wherein the enzymatic activity is measured using a cellobiose activity assay.

6. The composition of claim 1, wherein the variant Cel5a endoglucanase comprises a combination of point mutations in the wild type Cel5a amino acid sequence of SEQ ID NO: 4, the combination of point mutations being selected from the group of combinations consisting of:

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309L, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309W, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309L, and S318E;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309W, and S318E; and

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, F191V, T233V, G239E, V265T, D271Y, G293A, S309W, and S318P.

7. A method of hydrolyzing cellulose, comprising hydrolyzing the cellulose with the composition of claim 1.

8. The method of claim 7, wherein the hydrolyzing occurs at a temperature of about 60° C. to about 70° C.

9. A composition comprising a variant Cel5a endoglucanase having a temperature at which half of the maximal protein activity remains ( $T_{50}$ ) higher than a  $T_{50}$  of wild type Cel5a endoglucanase of SEQ ID NO: 3, the variant Cel5a endoglucanase comprising point mutations in a wild type Cel5a amino acid sequence of SEQ ID NO: 4 selected from the group consisting of

(a) G293A and S318P/Q/E and

(b) G293A and S318P/Q/E and one or more residues selected from the group consisting of E53D, T57N, S79P, T80E, V101I, S133R, N153D, N155E, T156E, G189S, F191V, T233V, G239E, V265T, D271Y, and S309W/L, in which the amino acid residue numbering is based on the numbering of (SEQ ID NO: 5).

10. The variant Cel5a endoglucanase composition of claim 9, wherein the point mutation comprises a combination of mutations selected from the group of combinations consisting of:

T57N, N76P, T80E, S139P, N155E, G189S, K219Q, G239N, D271F, Y278F, G293A, and S318P;

T57N, N76P, T80E, S139P, N155E, G189S, K219Q, G239N, D271F, Y278F, G293A, S309F, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309L, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309W, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309L, and S318E;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309W, and S318E; and

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, F191V, T233V, G239E, V265T, D271Y, G293A, S309W, and S318P.

11. A method of expressing a variant Cel5a endoglucanase protein having increased thermostability compared to wild type Cel5a endoglucanase of SEQ ID NO: 3, the method comprising expressing the composition of claim 9 in a host organism.

12. The method of claim 11, wherein the host organism belongs to a genus selected from the group consisting of *Saccharomyces*, *Pichia*, *Trichoderma*, and *Aspergillus*.

13. A composition comprising a variant Cel5a endoglucanase having increased protein expression in a host organism compared to wild type Cel5a endoglucanase of SEQ ID NO: 3, the variant Cel5a endoglucanase comprising point mutations in a wild type Cel5a amino acid sequence of SEQ ID NO: 4 selected from the group consisting of

(a) G293A and S318P/Q/E and

(b) G293A and S318P/Q/E and one or more residues selected from the group consisting of E53D, T57N, S79P, T80E, V101I, S133R, N153D, N155E, T156E, G189S, F191V, T233V, G239E, V265T, D271Y, and S309W/L, in which the amino acid residue numbering is based on the numbering of (SEQ ID NO: 5).

14. The variant Cel5a endoglucanase composition of claim 13 having increased protein expression, wherein the point mutations comprise a combination of mutations selected from the group of combinations consisting of:

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309L, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309W, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309L, and S318E;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309W, and S318E; and

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, F191V, T233V, G239E, V265T, D271Y, G293A, S309W, and S318P.

15. A composition comprising a variant Cel5a endoglucanase having a temperature at which half of the maximal protein activity remains ( $T_{50}$ ) higher than a  $T_{50}$  of wild type Cel5a endoglucanase of SEQ ID NO: 3, increased enzymatic activity relative to the wild type Cel5a endoglucanase of SEQ ID NO: 3, and increased protein expression in a host organism relative to the wild type Cel5a endoglucanase of SEQ ID NO: 3, the variant Cel5a endoglucanase having a combination of mutations in a wild type Cel5a amino acid sequence of SEQ ID NO: 4 in which the amino acid residue numbering is based on the numbering of (SEQ ID NO: 5), the combination of mutations being selected from the group of combinations consisting of:

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, G239E, D271Y, G293A, S309L, and S318P;

E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309W, and S318P;  
 E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 G239E, D271Y, G293A, S309L, and S318E;  
 E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S, 5  
 G239E, D271Y, G293A, S309W, and S318E; and  
 E53D, T57N, S79P, T80E, V101I, S133R, N155E, G189S,  
 F191V, T233V, G239E, V265T, D271Y, G293A,  
 S309W, and S318P.

**16.** The composition of claim **15**, wherein the host organ- 10  
 ism belongs to a genus selected from the group consisting of  
*Saccharomyces*, *Pichia*, *Trichoderma*, and *Aspergillus*.

**17.** The composition of claim **13**, wherein the variant Cel5a  
 endoglucanase further comprises a point mutation selected  
 from the group consisting of D13E, T18P, G64A, N76P, 15  
 S79P/Q/E, T80E/Q, I82L/M, V101I/L, A122E, S133R,  
 S139P, N155E/Q, T156E, G189S/A/E/K, K219Q/A, G239E/  
 N/D, V265T, D271Y/F, Y278F, and S309W/F/L.

\* \* \* \* \*